THE

LONDON, EDINBURGH, AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE.

[SIXTH SERIES.]

JULY 1923.

I. *The Theory of Mensuration.* *By* DOROTHY WRINCH, *D.Sc., Fellow of Girton College, Cambridge, and Member of the Research Staff, University College, London, and* HAROLD JEFFREYS, *M.A., D.Sc., Fellow of St. John's College, Cambridge* *.

PART I.—PRELIMINARY CONSIDERATIONS.

The Nature of Laws and Individual Experience.

THE interest of science rests principally in its laws, and not in its data. Particular sensations have interest for science only in so far as they make it possible to discover and verify laws; the experiencing of a sensation, apart from the conditions that accompany it, affords no basis for an inference, and must remain permanently part of the unsystematized data not included in science. The data that can be utilized are simultaneous occurrences of two or more sensations, not the individual sensations themselves; they are therefore instances of relations between sensations. The discovery of a law consists in finding a relation that is common to a considerable number of data. From such a relation, given that a certain set of sensations (which we shall call the circumstances) occurs, it is possible to infer that another sensation determined by the relation must also occur.

* Communicated by the Authors.

It is thus of fundamental importance to notice the difference in structure between a law and a proposition embodying the result of a single observation. We may denote a law, a proposition that a certain relation between circumstances k and sensations a will always be found to hold *, however k and a themselves may vary, by $(m)f(a_m, k_m)$. A particular datum or consequence of the law may then be denoted by $f(a_s, k_s)$. The interest for science of the fact that a certain investigator at a certain moment at a certain place had a certain set of sensations exists only in the possibility that it offers of affording an instance of a general proposition relating such sets of sensations.

The above notation draws attention to the fact that there is no essential structural difference between a law $(m)f(a_m, k_m)$ that is capable of verification, and one $(m)f(b_m, l_m)$ that is incapable of verification. In the one case it is asserted that a sensation a will be experienced under conditions that can in point of fact on occasion be satisfied. In the other case it is asserted that a sensation b will be experienced, given that certain conditions are satisfied, the conditions being such that they cannot in fact be satisfied. But there is no difference between the forms of the relations between the sensations and the circumstances in the two cases. In the case of Einstein's theory of gravitation, for instance, the tests of the motion of the perihelion of Mercury and the bending of a ray of light are instances of the first type of law; while it appears probable that the law of the shift of the lines in the solar spectrum could be tested only in circumstances that cannot, in fact, be realised.

There is no methodological difference between the data obtained by experiment and observation. In each case an association of sensations is recorded, and laws are inferred purely from such associations. The difference between experiment and observation is that in the latter the combination of circumstances k_m is not subject to the control of the investigator, who merely notes the occurrence of a_m at the instant when his changing sensations include k_m; whereas in an experiment the investigator brings about the conditions k_m himself. Usually k_m depends only partly on the experimenter's actions. Perhaps an experiment on cathode rays represents the extreme of control that is possible; while extreme lack of control might be represented by eye observations, without a nephoscope, of the

* The symbol $(x)f(x)$ is habitually used in modern logic to denote the assertion that, for all values of x, the proposition $f(x)$ is true.

direction of motion of clouds. Even in the latter case, however, action on the part of the observer is necessary to obtain any result: he must open his eyes and look at the sky. Most scientific observations, however, are intermediate in character. In astronomy, for instance, the motion of the heavenly body is independent of the observer, but the telescope must be entirely under his control for the results to be of any value. In every case, however, what matters is the fact that a certain sensation was experienced in conjunction with certain others; in the subsequent development one utilises this fact, and no further reference is made to the part played by the observer in bringing about the conditions necessary to make the observation possible.

In some cases, laws will differ only in that the circumstances k_m under which a certain sensation is asserted to be experienced are replaced by k_m', where k_m' is more or less narrowly specified than k_m. In general, the specification of a will not be enough to determine the conditions of its occurrence, though the truth of a law asserts that the circumstances entail the occurrence of a.

At any one stage of scientific knowledge, the important question is whether certain consequences of the laws under consideration at the moment will be verified or not. In the case of a law with several verifiable consequences, some of which have already been verified, we have to assess the probabilities of certain other propositions, given these various pieces of information. We have, in fact, to deal with

$$P\{f(a_m, k_m) : f(a_1, k_1) \cdot f(a_2, k_2) \ldots f(a_{m-1}, k_{m-1}) \cdot h\}$$

if we wish to discuss the probability that some further prediction will be verified, and with

$$P\{(m)f(a_m, k_m) : f(a_1, k_1) \ldots f(a_s, k_s) \cdot h\}$$

if we are considering the probability of the law itself.

The Modified Simplicity Postulate.

It is of vital importance for the understanding of the propositions of science that one should realize that the term k occurs in every proposition. Hence the simplicity postulate introduced in a previous paper requires a slight modification, since it evidently needs to be stated in a form that

makes the k term explicit. We therefore restate it in the following form :—

If $(m)f_1(a_m, k_m)$, $(m)f_2(a_m, k_m), \ldots (m)f_n(a_m, k_m)$ are all general laws, concerning the truth of which for the admissible values of k we have no previous relevant information, the prior probabilities of these laws, namely

$$P\{(m)f_1(a_m, k_m) : h\},$$
$$P\{(m)f_2(a_m, k_m) : h\} \ldots P\{(m)f_s(a_m, k_m) : h\}$$

are all finite, but constitute the terms of a convergent series, and accordingly are $\aleph_0$ in number ; further, the simpler the law, the higher its prior probability.

By the method of our former paper we have

$$\begin{aligned} &P\{f(a_1, k_1) \cdot f(a_2, k_2) \ldots f(a_m, k_m) : h\} \\ &= P\{f(a_m, k_m) : f(a_1, k_1) \ldots f(a_{m-1}, k_{m-1}) \cdot h\} \\ &\qquad \times P\{f(a_1, k_1) \ldots f(a_{m-1}, k_{m-1}) : h\} \\ &= P\{f(a_1, k_1) : h\} \, P\{f(a_2, k_2) : f(a_1, k_1) \cdot h\} \ldots \\ &\qquad \times P\{f(a_m, k_m) : f(a_1, k_1) \ldots f(a_{m-1}, k_{m-1}) \cdot h\}. \end{aligned}$$

Our hypothesis requires that the quantity on the left shall be finite, and shall tend to a definite limit different from zero as m tends to infinity ; for its limit is the probability of the general law $P\{(m)f(a_m, k_m) : h\}$. Hence, whatever number λ less than unity is selected, there cannot be more than a finite number of factors less than λ on the right. Hence, as m increases,

$$P\{f(a_m, k_m) : f(a_1, k_1) \ldots f(a_{m-1}, k_{m-1}) \cdot h\}$$

must tend to unity as a limit. But $f(a_m, k_m)$ is the proposition that a_m will be experienced if k_m occurs; and therefore it follows that the probability that the next verification will succeed if tried approaches to certainty as the number of verifications increases.

Note.—Prof. G. H. Hardy has called our attention to a slip in a previous paper (Phil. Mag. xlv. p. 370, 1923). A law involving a constant differing from an integer by $\frac{1}{n}$, where n is a large integer, would be expected to occupy a position in the probability series later than the mth, where

$$m = \sum_1^n \phi(n) = \Phi(n),$$

where $\phi(n)$ is the number of numbers less than n and prime to it. The asymptotic value of $\Phi(n)$ is $\frac{3x^2}{\pi^2} + O(x \log x)$, given in Landau's *Primzahlen*, p. 579.

The Verification of Intermediate Steps.

The argument used in our former paper, concerning cases where the success of repeated verifications may make the law itself have a probability approaching certainty when the verifications are included among the relevant data, is still applicable when the constituent propositions are of the form $f(a_s, k_s)$. Consider now the law $(m)f(a_m, k_m)$, and its posterior probability based on several data of the form $f(a_s, k_s)$. These data may arise in two ways: first, they may be derived directly from experience; second, they may be inferences from laws already known. Considering the second possibility first, we notice that such inferences may themselves be laws of considerable generality; they may have probabilities approaching certainty, on account of the verifications of the laws they are derived from. When they are to be used in further inferences, it is quite unnecessary to verify them directly by experiment, for they are used only in the form "if k occurs, a will also occur"; and if this datum is practically certain from previous knowledge, it is a matter of complete indifference to its subsequent application whether it is verified directly. Indeed, it does not matter whether it is possible to verify it at all. Thus we arrive at the important general principle that the verification of intermediate steps is unnecessary. This will be used repeatedly in what follows; just as it has been used implicitly in every work on theoretical physics ever written.

In the case of a proposition $f(a, k)$ derived directly from experience, we evidently do not require to insert as a separate datum in the probability estimate the fact that the circumstances k occurred, since this fact does not affect the probability of the propositions under consideration. For, first, if any k has not occurred, the corresponding a has not been verified as occurring with it, and therefore $f(a, k)$ is not part of the available data relative to the law we are considering. Secondly, if it has occurred, it conveys no relevant information without the corresponding a, and therefore its mention except in its association with a is unnecessary.

Judgments of Irrelevancy.

The knowledge available with respect to any experiment or observation has reference to a very large number of circumstances. Thus, in the account of the experiment we may know the temperature of the laboratory, the time of day, the type of telescope used, and many other details.

The characterization will not be a complete account of the conditions of the experiment unless these details are exhaustively enumerated. In order to relate to a law a fact whose specification we have, it is customary to take into account only a few of them, and it is considered that the others are irrelevant to the purpose in hand.

It will be convenient to use k to denote only the aggregate of relevant circumstances, excluding from it the other circumstances involved. To say that the datum b is irrelevant is then to say that, for all values of b, $P\{(f(a, kb) : h\}$ is the same, provided that k remains unchanged. Thus to assert that b is irrelevant to a set of sensations is to say that the same sensations a occur with one value of b as with another. An assertion of irrelevance of this type is capable of direct test. For suppose a number of tests applied in which b has varying values, and that a is found to remain the same. There are two alternatives. First, a may really depend on both b and some other datum c, and the change in a dependent on the variation of b is just balanced by that due to the variation of c. This has an infinitesimal prior probability, since for every value of b it requires c to have a definite value; unless indeed b and c are themselves connected by a law, when the connexion can be observed and the law discovered, and it then ceases to be necessary to specify some part of bc. The alternative is to suppose that a is, in fact, independent of b. But independence is the simplest of all laws, and therefore always has a considerable prior probability. It corresponds to the simplest possible differential equation, namely

$$\frac{dy}{dx} = 0.$$

The two hypotheses, (1) that a and b are, in fact, independent, and (2) that the effect on a of variations in b is just annulled by unrelated variations in c, both imply the lack of experimental evidence for a connexion. The posterior probabilities, given that no connexion has been found between a and b, are therefore in the ratio of the prior probabilities, and therefore it is practically certain that a and b are independent. Thus the experimental establishment of independence is possible, and it is probable that independence is usually an empirical inference and not an *a priori* hypothesis. An electrical experiment is expected to give the same result independently of the time of day and the position of the laboratory, not because one has a prior belief that either

absolute or relative position and time are irrelevant, but because it was actually empirically proved that they are irrelevant in the course of the experiments that established the electrical law.

Part II.—The Theory of Mensuration.

Mensuration may be defined as the science of the relations between measurements of distance in rigid bodies. It must be carefully distinguished from all forms of geometry, on account of the utter dissimilarity of the methods of development. The method of geometry is essentially intensive, while mensuration is essentially extensive, involving a principle of empirical generalization, like all other empirical sciences *. Each geometry rests on a number of *a priori* general postulates. The development of any one geometry consists in the deduction of the consequences of any one set of mutually consistent postulates. In mensuration, on the other hand, we have to start with results obtained by actual experiment, which by their very nature have been tested for only a finite number of instances, and the results covering all cases, or indeed any other cases, arise only as generalizations in the course of the development, and not as part of the preliminary postulates.

It may be suggested that before a discussion of mensuration can be undertaken, the terms "distances," "rigid bodies," which are used in our specification of the subject-matter, should be defined. The reply to this requires a digression. In accordance with the principles enunciated in a former paper †, the main requirement of a definition is to make it possible to recognize the defined notion when it actually occurs. It is of no value to say that a rigid body is one such that the distances between all the points of it are unaltered by any displacement, nor to define relative motion as change of distance between parts of the system, unless we have some way of recognizing when distances *are* altered. Distance, again, cannot be defined in terms of the properties of rigid bodies unless we have first some method of recognizing the rigid body when we meet it. None of these notions can be defined in terms of the properties of space, because we have no means of recognizing space directly; distance in space, for instance, cannot be determined except through measurements, which at once re-introduce

* For a fuller discussion of the distinction, see 'Nature,' Feb. 27, 1921, pp. 806–809.

† Phil. Mag. vol. xlii. pp. 369–390 (1921).

material scales, which the reference to space was intended to avoid.

The solution appears to be that none of these notions is directly recognizable, and that all are derived from still more elementary notions, several experimental facts being used in the process. Let us start from the notion of a body, without considering the processes by which we arrive at this. It is a fact that we can make permanent marks on bodies, which we can recognize afterwards. It is also a fact that two bodies can be made with "edges," so that when they touch at two parts of an edge they touch at all intermediate parts; also however they may be turned, subject to these two points remaining in contact, the intermediate parts all remain in contact. Given that this has been observed a sufficient number of times with one pair of bodies, we can use our principle of empirical generalization to infer with a high degree of probability that it will be the case in any subsequent trial with these bodies. In such cases we may call the edges "straight." The reservation must be made that the bodies must receive only ordinary treatment during the test. It is easy to recognize by our sensations of force when exceptional treatment is taking place. If edges fail to satisfy the test, they do not form part of the subject-matter of mensuration; and if edges that have been found to satisfy it in many previous trials fail in a new trial, we say that the conditions in the latter were exceptional. Thus experiment enables us to classify edges into (1) those straight and suitable for mensuration, and (2) those never straight or not permanently straight under ordinary conditions. The occurrence of failure in similar experiments on edges previously regarded as straight provides a standard of what conditions are to be regarded as exceptional. The same type of experiment therefore leads to standards both of straight edges and of ordinary conditions.

The straight edge may now be used to construct a definition of a rigid body. Marks may be made at intervals, quite arbitrarily, along the edge, and may also be made on the surface of a body. It may be found that when we put a certain mark on the edge in contact with a certain mark on the body, the scale can be turned so that another mark on the body lies between two other consecutive marks on the edge. If now, no matter how the edge and the other body may be turned, this mark always lies between the same two marks on the scale, we may infer that this will be true in any subsequent trial. If the same is found to be true in subsequent trials with different marks, the body may

be called rigid, and the straight edge also may be said to be "on a rigid body."

An important method of testing "rigidity" is by means of such instruments as calipers or compasses. These being set with one point at a mark on the body, the other point may be made to mark out a curve on it. If in subsequent trials it is always found that when one point of the compasses is at one mark the other, if in contact with the body at all, touches it at some point on the curve already drawn, the body may be called "rigid." The compasses or calipers themselves, between adjustments, may be shown to be rigid bodies by comparison with a straight edge. Their use is indispensable in testing the rigidity of convex solids. Cases where the rigid body tests are not satisfied are not included in the subject matter of mensuration, but are dealt with in higher branches of mechanics.

We next require a rule for ordering distances. If we have two pairs of marks, A and B on one body, C and D on another, and apply the caliper test to them, we may find that if the calipers are set so that one point can rest on A and the other on B at the same time, and if one point is then applied to C and a curve marked out on the body with the other point, D is either within, on, or outside this curve. If it is within it, we agree to say that the distance AB is greater than CD; if it is on it, that they are equal; and if it is outside it, that AB is less than CD. The process may be inverted. We can establish experimentally in this way that $AB=BA$; that if AB is greater than CD, then CD is less than AB; and that the relations greater than, equal to, and less than, are transitive. Thus these three propositions, which are postulates in most metrical geometries, are experimental facts in mensuration. We have then arrived at a definition of what is meant by saying of two distances that one is greater, less than, or equal to the other.

In the above account the straight edge has been treated as anterior to the rigid body, whereas it might appear that the opposite procedure would have been the more natural. The arrangement which has been adopted here appears to be necessary, on account of a difficulty in the comparison of distances. In order to establish directly that the distance associated with one pair of marks *ab* is less than the distance associated with another *cd*, it is necessary to place the bodies bearing them in such a position that one can observe that neither *a* or *b* is outside *cd*. This can be done only if at least one of the pairs *ab*, *cd* is connected by a straight

edge; for otherwise the marks may only be missing each other as the bodies are turned about, without a or b ever being between c and d at all.

The Numbering of Distances.

The facts of experience and the principle of empirical generalization therefore enable us to build up a theory of mensuration from the primitive idea and a few simple postulates. We may now go on to consider the way in which more complicated results may be obtained. We have so far considered only the relations greater than, less than, and equal to between distances on rigid bodies associated with marks. We now propose to build up a quantitative system of measurement such as is used in scientific instruments.

If we take any two straight edges, we can make marks along them when in contact so that to every mark on each there corresponds one on the other in contact with it. The results already enunciated are enough to establish the proposition that if the distance between two points on a rigid body is between the distances from the first mark on one of these edges to the nth and the $(n+1)$th respectively, it will also lie between the distances on the other scale that corresponded to these in the test. Further, it is not necessary that scales should have been directly compared. It is sufficient that they should have been compared to the same scale. Hence, if we have one standard scale, it is possible to make others to its pattern, and by means of them to assign an order to all distances sufficiently small to come within the scope of any of them. Similarly, the caliper method may be used to compare distances on a body with those on a scale, the calipers when set to one adjustment being a rigid body within the definition.

So far the scales considered have been graduated arbitrarily, and the results apply equally to all methods of graduation. There is one special type that offers great advantages over all others. If two distances a and b are set off along a straight edge from the point o, a definite mark is reached; this is experimentally verifiable. If the same two distances are set out in the reverse order, b first and then a, it is found that the same mark is reached. This is a new result, and is not a consequence of our previous ones. Repeated experimental confirmation, however, enables us to attach practical certainty to inferences from it. It is also found that however many distances are set out along a

straight edge in this way, the order in which they are set out makes no difference to the point where we finish. These facts suggest a relationship to the numerical process of addition. We can make marks along a straight edge so that every distance from a mark to the next is the same, except of course in the case of the last one, where there is no such distance. Such a marked edge may be called a uniformly graduated scale. The measured distance from one point to another can now be defined as the number of such intervals along this edge that it can be made to overlap, when compared with the edge by rigid body displacements. The measured length of a segment of a straight edge may also be defined as the measured distance between its ends. In consequence of the results of this paragraph, the measured length of any segment of a straight edge is equal to the sum of the lengths of any segments into which it may be subdivided (with slight errors, small in comparison with the whole length when the measured length is great). Thus numerical addition is applicable to measured lengths.

The measuring scales used in laboratories are just such scales as we have been describing. In the process of manufacture the successive divisions are marked off automatically, so that the marks can be made to touch simultaneously consecutive turns of a screw, which is a rigid body according to the definition adopted in this paper. Further, this can be done whatever mark we start with, so that the screw serves also as our intermediate body for the comparison of one segment of the straight edge with another, and the length of every segment is thus found to be the same. The process of manufacture of the screw is itself such that every turn is compared with a standard body, to which all the intervals on the scale are therefore ultimately referable. Thus laboratory scales are uniformly graduated scales, and measures made with them come within the subject matter already marked out.

A definition of relative motion is not required at this stage. We are dealing only with cases where the measured distances are found, on repetition of the process of measurement, not to vary. Hence the absence of relative motion among the parts of a system has already been included in the specification of the problem.

It may be objected that there are many bodies which do not satisfy the conditions we have specified. It is, however, an experimental fact that there are many bodies which do satisfy them. These bodies form the subject matter of the

theory of mensuration. Other bodies which do not satisfy these conditions obviously exist, but their treatment is a matter for other branches of physics.

In many cases it may turn out that a distance whose value has been inferred according to the rules which have been enunciated above cannot actually be measured, on account of the impossibility of carrying out some physical construction. This affords no objection to the methods in question. If it is possible to carry out the construction, the methods here elaborated predict what the measure will be. The physical construction required—if indeed any is required—may be regarded as part of the perceiving arrangement, and experiments which cannot be carried out as on a similar footing to sensations which cannot be perceived; they are not part of the subject matter of science.

Can Mensuration be included in any known Geometry?

While the development of mensuration differs from that of any existing geometry in its use of a principle of generalization, it may yet happen that when its elementary results have been generalized by this principle the generalizations may be found to agree with those of some known geometry. If this were so, the whole of it would constitute a class of entities and relations to which the body of propositions that constitute this geometry would be applicable. The only reservation to be made would be the one * about inferred values lying within certain limits and not necessarily being exact. When we examine geometries, however, we find that there is none that satisfies our conditions. Consequently the whole of the theory of mensuration must be developed from the beginning.

All projective and descriptive geometries are evidently ruled out at once. A requirement of all such geometries is that no notion analogous to distance is to be used. Since distance forms our subject matter, there is no common ground whatever.

Euclid's geometry is the closest existing analogue of mensuration, and a more detailed discussion of it is desirable. The notion of length is freely used in it, being treated as undefined. His points are sufficiently like our marks, and his straight lines are sufficiently like our straight edges. His unstated axiom of superposition is practically what we have adopted as an experimental fact for rigid bodies.

Nevertheless Euclid's geometry is not applicable to

* Phil. Mag. vol. xlv. pp. 368-374 (1923).

mensuration. His "length" is applied to curves as well as to straight lines (see Prop. VI. 33), which we have not yet had occasion to do; but in fact most of his applications are to straight lines. He postulates that every straight line has a length. The corresponding proposition in mensuration is true; for though we can deal only with measures of length, and it happens sometimes that a straight edge is longer than any available measuring scale, we can always repeat the application and by an analogue of the well-known postulate of Archimedes continue until the whole of the edge has been overlapped, and a measure is then obtainable. But he postulates further that any two points can be joined by a straight line. The analogue of this is often true, but in the case of a convex body too hard to be bored it is not possible to construct a straight edge that extends from one to the other. Yet the distance between these marks exists, for it can be measured by the caliper method. Thus Euclid's treatment fails to satisfy the condition stated in a previous paper* to be essential to any scientific theory—namely that of being applicable in practice to the subject matter with which it is proposed to deal.

The most important departure of Euclid's treatment from any possible account of mensuration, however, is in the discussion of parallels and the related propositions. We refer especially to the fifth postulate, also called the twelfth axiom, and to the second postulate, that a straight line may be produced to any length, however great. Both of these postulates have been criticized by modern geometers as not obvious. In mensuration, on the other hand, they are not only not obvious but demonstrably false. The length to which we can produce a straight edge is limited by the size of the body of which it forms part; it may be extended by fastening other bodies on, but there is a limit to this process, and therefore to the length of a straight edge. Again, it may be possible to find out by our existing methods of measuring angles that, when one straight edge crosses two others, it makes the sum of the interior angles less than two right angles, but it does not happen in all such cases that the two straight edges it crosses intersect: for in fact they are often too short, or they may not be in one plane—a detail that is not allowed for in the usual statements of the postulate.

The alternative known as Playfair's axiom does not meet

* Phil. Mag. vol. xlii. pp. 369-390 (1921).

the difficulty, for it is not true that of two intersecting straight edges at least one must intersect any other: Playfair's parallel axiom fails in just the same way as Euclid's.

An alternative method would be to adopt as an experimental fact, which it may seem to many to be, that the sum of the angles of a triangle is equal to two right angles. We do not wish to deny that an account of mensuration may be constructed on these lines; but we do not think it is the best way of constructing one. So far we have not had occasion to specify how angles are to be measured. Euclid gives no method of measuring them, but assumes in the course of his proofs several propositions about them that amount to such a method. He supposes in I. 4 that angles that can be superposed are equal; and in I. 13 that if a pencil of coplanar lines is drawn through a point, the angle between the extreme lines is equal to the sum of those between consecutive lines of the pencil. In other words, he supposes that there is a quantity associated with any pair of lines which is the same for superposable pairs on rigid bodies, and is additive for sheaves of angles in a plane. This can be proved experimentally in many cases by means of the protractor.

This method is, however, of very limited application in practice. The angles can only be superposed in special cases; projections on the rigid bodies that carry them usually interfere with it. Again, the truth of the addition proposition rests on the angles to be added being placed in the same plane; thus the measurement of angles is posterior in knowledge to the notion of a plane, and we require a rule that will enable us to recognize a plane when we meet one. Euclid's definition is that a surface is plane if any straight line joining two points of it lies wholly upon it. But, in the case we are thinking of, projections will again nearly always interfere with the application of this test. Hence the notion of a plane, and with it that of the direct measurement of angles, should be avoided in a development of mensuration if it can be managed.

In a paper recently published * a very different view of Euclid is taken. It is however admitted (p. 28) that several of Euclid's postulates are experimentally false.

* Campbell, Phil. Mag. vol. xliv. pp. 15–29 (1923).

The Development of a Theory of Mensuration.

In our discussion of mensuration we have, so far, arrived only at methods of comparing distances, estimating them quantitatively by comparison with uniformly graduated scales, and at propositions about distances along the same straight edge. The way in which real numbers may be applied to the distances associated with pairs of points may be exhibited in the following postulates and definitions :—

1. To each pair having a distance, one and only one number belongs.
2. If the number n belongs to the pair (x, y), then it belongs also to the pair (y, x).
3. If the distance between one pair of points is not greater than that between another pair, then the number belonging to the first is not greater than the number belonging to the second.
4. If two pairs (x, y), (y, z) are such that $x\,y\,z$ are in the same straight line (with y between x and z), then the number belonging to the pair (x, z) is the sum of the numbers belonging to the distances associated with the pairs (x, y), (y, z).
5. The number 0 is to belong to the distances associated with pairs in which the two constituents are identical.
6. The number 1 is to belong to the distance associated with a certain pair called the unit pair.

Thus any one system in the theory we are suggesting would involve these postulates, together with some arbitrary choice of the unit pair defined in 6. Thus the Metric system and the British system differ only in the fact that in the arbitrary choice of the unit pairs the unit pair selected in the one system was not congruent to the unit pair selected in the other system. Thus, if the actual marks corresponding to these unit distances were placed side by side, the marks would not be superposable.

With these postulates and definitions, we write for the number belonging to the pair (x, y) the symbol xy. Thus we have

$$xy = yx,$$

$$xx = 0.$$

It is worth while to point out that the assertion

$$xy = 1$$

means that the pair (x, y) is the actual pair chosen as the

pair to have unit distance associated with it or congruent to it.

When the postulates governing the number of distances have been specified, it is permissible to use all algebraic and arithmetical operations. It will therefore not be necessary to define, for example, what is meant by the square of a number *. Arithmetic and Algebra are taken over in their entirety.

The account which was given on p. 9 of the development of the idea of distance arose directly out of notions which can evidently be reduced to the one primitive notion of "not greater than." In introducing measurement, we have had to introduce a further notion, namely that of a fixed standard of reference. This standard could have been chosen in many ways from the pairs available, and it is not as a rule specified by previous experience. If a different standard is chosen, different numbers are found. Thus our measures are apparently influenced by convention. It can readily be inferred from the results already stated, or it may be directly verified, that a change of standard alters all the numbers in measures in the same ratio. Accordingly, the conventional choice of a unit of distance does not influence the ratios of the numbers involved in measures of distance, and therefore if we discover any physical law homogeneous in these numbers that is true with one unit, it will be true whatever other unit is chosen. We shall, in fact, use only such laws in the development of the subject. It could not have been predicted beforehand that any law would exist that would involve only homogeneous functions, so that the unimportance of convention is in this case, as in all others, a matter of experimental verification rather than of *a priori* necessity. On the other hand, in inferences from laws it will be necessary to obtain the actual measure in terms of some particular unit; in this case the unit must of course always be specified, for otherwise the result would be meaningless.

We also require propositions about the relations between the distances from each other of marks not in the same straight edge. This introduces a new domain, and we need at least one new experimental fact to serve as a starting point. As has already been indicated, propositions involving angles or planes should be avoided if possible. We therefore need an experimental fact dealing

* This and other relevant definitions are given in 'Principia Mathematica,' Whitehead and Russell, Cambridge. *Cp.* *91, where the square of a relation is defined.

with distances. The most suitable appears to be the following:—

Let two straight edges meet in o. Let x be a mark on one edge, and y a mark on the other. Then, whatever ox and oy may be,

$$\frac{\partial}{\partial ox}\left(\frac{ox^2+oy^2-xy^2}{2ox \cdot oy}\right)=0,$$

$$\frac{\partial}{\partial oy}\left(\frac{ox^2+oy^2-xy^2}{2ox \cdot oy}\right)=0,$$

and each ratio is not greater in absolute value than 1.

This proposition lacks the chief qualification of a primitive proposition in a geometry—namely, that of possessing a *naïreté* that disarms suspicion. For our purpose, on the other hand, it has the great advantages that it is capable of experimental test in almost all cases, and that such test has already been carried out in countless experiments in practical plane "geometry." So far as we know, it has not been tested with the full accuracy of which modern measuring apparatus is capable, but enough has been done to establish it in an enormous number of cases. It is not extremely simple in form, but the number of verifications is so great that if its prior probability is at all appreciable, the probability of all inferences from it must amount to practical certainty. We therefore suppose it to hold in general, and attempt to develop its consequences.

Instead of working with the ratio itself, it is often convenient to use a certain known function of it. If the straight edges oa, ob meet in o, we put

$$\frac{oa^2+ob^2-ab^2}{2oa \cdot ob}=x.$$

We can now define the measure of the angle aob to mean $\int_x^1 \frac{du}{(1-u^2)^{\frac{1}{2}}}$. This is a known function denoted in pure mathematics by $\cos^{-1} x$. The path of integration is confined to real values of u, and the positive sign is taken for the root, so that no difficulty of interpretation arises. Then we have

$$\cos aob = x,$$
$$\sin aob = +(1-x^2)^{\frac{1}{2}}.$$

It is evident from our premises that the measure of an angle is the same whatever marks on its arms we use in our length-measurements.

1. If two straight edges meet, the adjacent angles are together equal to π.

Let a be a mark on one edge, and b_1 and b_2 marks on the other, such that o the junction of the edges lies between them. We have

$$ab_1^2 = oa^2 + ob_1^2 - 2oa \,.\, ob_1 \cos aob_1, \quad \ldots \quad (1)$$

$$ab_2^2 = oa^2 + ob_2^2 - 2oa \,.\, ob_2 \cos aob_2, \quad \ldots \quad (2)$$

and also $$= ab_1^2 + b_1b_2^2 - 2ab_1 \,.\, b_1b_2 \cos ab_1b_2. \quad \ldots \quad (3)$$

Now $$b_1b_2 = b_1o + ob_2, \quad \ldots \quad (4)$$

$$\cos ab_1b_2 = (ab_1^2 + ob_1^2 - ao^2)/2ab_1 \,.\, ob_1$$
$$= (ob_1 - oa \cos aob_1)/ab_1 \quad \ldots \quad (5)$$

by (1).

Substituting in (3) from (1), (4), and (5), we have

$$ab_2^2 = oa^2 + ob_2^2 + 2oa \,.\, ob_2 \cos aob_1. \quad \ldots \quad (6)$$

Thus by comparison with (2) we find

$$\cos aob_2 = -\cos aob_1;$$

whence $$aob_1 + aob_2 = \pi. \quad \ldots \quad (7)$$

2. It follows as an immediate corollary by Euclid's method that when two straight edges cross, the opposite angles are equal.

3. The sum of the angles of a triangle is equal to π.

Consider any triangle abc. Put $bc + ca + ac = 2s$.

Then, defining the tangent as the ratio of sine to cosine and transforming as in trigonometry, we find

$$\tan \tfrac{1}{2} bac = \left\{ \frac{(s-ab)(s-ac)}{s(s-bc)} \right\}^{\frac{1}{2}},$$

with two symmetrical relations. Hence, by the properties of the tangent, we can find

$$\tan \tfrac{1}{2}(bac + bca) = \left\{ \frac{(s-ca)}{(s-ab)(s-bc)} \right\}^{\frac{1}{2}}$$
$$= \tan \tfrac{1}{2}(\pi - abc).$$

Whence $$bac + abc + bca = \pi.$$

4. If two straight edges oa and ob meet at o, and if $\cos aob$ is negative, and if o is not the end of ob, it is possible to make a mark b_1 on ob such that $\cos aob_1$ is positive.

For by Prop. 1 we need only take b_1 on the side of o opposite to b, when the result follows.

5. If a be outside the edge ob, and if b be on that part

of it for which aob is less than $\frac{1}{2}\pi$, and if ob be greater than $oa \cos aob$, then it is possible to make a mark c on ob such that $oa^2 = oc^2 + ac^2$.

For we can make a mark c on ob at a distance $oa \cos aob$ from o. Then

$$\begin{aligned} ac^2 &= oa^2 + oc^2 - 2oa \,.\, oc \cos aob \\ &= oa^2 - oc^2, \end{aligned}$$

which proves the proposition.

We can now introduce the definition of perpendicularity. If the angle between two straight edges is $\frac{1}{2}\pi$, they are said to be perpendicular. DF.

If oa is a straight edge, with a mark c on it so that the angle bco is $\frac{1}{2}\pi$, where b is some mark not on the edge, the mark c is called the foot of the perpendicular from b to the edge. DF.

6. If two straight edges oa, ob intersect at o, and c is any mark in one of them; if also the length of the other exceeds $oc \sec aob$, then we can make a mark d on it so that c is the foot of the perpendicular from d to the first line.

For we can make a mark d such that $od = oc \sec aob$, and the perpendicularity follows by the method of Prop. 5.

7. If the angle aob is $\frac{1}{2}\pi$, we have at once

$$\begin{aligned} ob &= ab \cos abo, \\ oa &= ab \sin abo, \\ oa &= ob \tan abo, \end{aligned}$$

with corresponding formulæ for the other trigonometric functions. These results thus emerge as laws, and not as definitions as in ordinary discussions.

8. Consider three edges meeting together in o. It is always possible to fix a in one of them so that a is the foot of the perpendiculars from points b and c on the other two, since the condition of Prop. 6 can always be fulfilled by making oa short enough.

Then

$$ab = oa \tan aob, \quad \ldots \ldots \quad (1)$$

$$ob = oa \sec aob, \quad \ldots \ldots \quad (2)$$

$$ac = oa \tan aoc, \quad \ldots \ldots \quad (3)$$

$$oc = oa \sec aoc, \quad \ldots \ldots \quad (4)$$

$$\begin{aligned} bc^2 &= ob^2 + oc^2 - 2oa \,.\, oc \cos boc \\ &= oa^2(\sec^2 aob + \sec^2 aoc - 2 \sec aob \sec aoc \cos boc). \quad (5) \end{aligned}$$

Also

$$bc^2 = ab^2 + ac^2 - 2ab \, . ac \cos bac$$
$$= oa^2(\tan^2 aob + \tan^2 aoc - 2 \tan aob \tan aoc \cos bac). \quad (6)$$

Equating these two expressions, and multiplying by $\cos aob \cos aoc$, we find that

$$\cos boc = \cos aob \cos aoc + \sin aob \sin aoc \cos bac.$$

This is the analogue in mensuration of the well-known formula of spherical trigonometry.

It follows as a corollary that bac is independent of oa.

9. If bac forms a straight edge, $\cos bac$ is -1, and we have

$$\cos boc = \cos aob \cos aoc - \sin aob \sin aoc$$
$$= \cos (aob + aoc);$$

whence the measure of boc is the sum of those of boa and boc if their sum does not exceed π; if it exceeds π,

$$boc + aob + aoc = 2\pi.$$

We have thus proved that angles can be added if the same straight edge intersects their arms: this proposition, which is fundamental in Euclid's treatment, emerges here as a special case of Prop. 8.

If three straight edges have a mark in common and a fourth straight edge not through this mark intersects all three, they are said to lie in one plane. Df.

Any mark that a straight edge can pass through and intersect two other straight edges is said to be in the same plane as these straight edges. Df.

The angle between two edges both perpendicular to the same edge is called the angle between the planes containing this edge and the first two. Df.

It will be noticed that we have not defined the term "plane" as such, but only the expressions "lie in one plane," "in the same plane," and "angle between the planes." Thus we can attach meanings to these terms even though no physical plane has been constructed.

10. Consider any two points l and m, and a straight edge op. Suppose points a on op, b on ol, c on om to have been found, such that ba, ca are perpendicular to op. Let $ol = r$, $om = r'$, $\angle lop = \theta$, $\angle mop = \theta'$. Then

$$lm^2 = r^2 + r'^2 - 2rr' \cos lom$$
$$= r^2 + r'^2 - 2rr' (\cos \theta \cos \theta' + \sin \theta \sin \theta' \cos bac). \quad (1)$$

If ak be any other straight edge through a, such that $kabc$ are in a plane, let angle $kab = \phi$, $kac = \phi'$. Then

$$bac = \phi - \phi'.$$

Hence

$$\begin{aligned} lm^2 &= r^2 + r'^2 - 2rr'\{\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi-\phi')\} \\ &= (r\cos\theta - r'\cos\theta')^2 + (r\sin\theta\cos\phi - r'\sin\theta'\cos\phi')^2 \\ &\quad + (r\sin\theta\sin\phi - r'\sin\theta'\sin\phi')^2. \quad . \quad . \quad (2) \end{aligned}$$

We can regard oa as our initial line, and oa and ak as lying in our initial plane. Then (1) is the standard formula for the distance between two points in spherical polar coordinates.

We can also define $x\ y\ z$ as follows :—

$$x = r\sin\theta\cos\phi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$y = r\sin\theta\sin\phi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$z = r\cos\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Then

$$lm^2 = (x-x')^2 + (y-y')^2 + (z-z')^2. \quad . \quad . \quad (6)$$

The distance has thus been expressed in the standard form appropriate to Cartesian coordinates.

The above definition of Cartesian coordinates is applicable in all cases where it is possible to find the distances and bearings of our points, whereas the usual definition is not applicable except where we can actually find the perpendiculars from the points to the three coordinate planes. We still have to show that x, y, and z are actually Cartesian coordinates according to the usual definition on those occasions when this can be applied.

11. If a straight edge be perpendicular to two straight edges that meet it, it is perpendicular to any other straight edge that meets it in the plane of these two. The proof of Euc. XI. 4 is readily adapted to prove this. In such a case the first edge may be said to be perpendicular to the plane.

12. Consider any marks p, q with spherical coordinates (r_1, θ_1, ϕ_1) (r_2, θ_2, ϕ_2). The condition that op, oq shall be at right angles is that

$$pq^2 = op^2 + oq^2.$$

Thus

$$2r_1r_2\{\cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2\cos(\phi_1-\phi_2)\} = 0,$$

or

$$x_1x_2 + y_1y_2 + z_1z_2 = 0.$$

13. At any point on the initial line, y and z are clearly zero. Hence at any point in the plane through o perpendicular to this line we must have $z=0$.

Let p be any point (x_1, y_1, z_1) and q any point $(x_2, y_2, 0)$ in this plane. Then

$$pq^2 = (x_1-x_2)^2+(y_1-y_2)^2+z_1^2.$$

Evidently this is least for a given position of p if

$$x_2=x_1 \quad \text{and} \quad y_2=y_1.$$

Then we see that the shortest distance of p from the plane $z=0$ is z. Similarly we can show that when $\phi=0$ we have a plane such that y is 0 for all points on it, and that the shortest distance of p from this is y_1; and similarly for z. Thus the Cartesian coordinates are the shortest distances of the point considered from three fixed perpendicular planes.

Since the shortest distance of p from any point on the plane must be at least as short as the shortest distance from the point to any line in the plane, it follows that the shortest straight edge that can extend from p to a plane must also be the shortest distance of p from every straight edge in the plane that meets it, and therefore, by Prop. 5 and the definition of perpendicularity, must be perpendicular to every such straight edge, or, by the definition under Prop. 11, to the plane.

Thus the Cartesian coordinates have been identified with those defined in the usual way.

II. *The Absorption produced by Electrically Luminescent Sodium Vapour.* *By* F. H. Newman, *D.Sc., A.R.C.S., Professor of Physics, University College, Exeter* *.

[Plate I.]

1. *Introduction.*

Although the Bohr theory of the atom has not yet been developed to give a complete representation of the spectra of elements other than those of hydrogen and helium, it gives a general indication of their structure. The difficulties involved in their calculation have not been overcome at present. The 1·5 S ring of an atom represents the stable orbit, and all the lines contained in the series 1·5 S$-mp$ are emitted when the electrons in different atoms fall from one of the p rings to the 1·5 S ring. The first subordinate

* Communicated by the Author.

series lines are produced by electrons falling from the *md* rings to the $2p$ ring, and the second subordinate series when the electrons fall from the *ms* rings to the $2p$ ring. Inter-orbital motions other than those mentioned give rise to combination lines and the fundamental series. The 1·5 S ring represents the outermost stable orbit in the normal atom, and the innermost unstable orbit. In order that the electron shall be ejected from the 1·5 S ring to the next orbit—the $2p$ ring,—the atom must receive energy either by bombardment, or by the absorption of radiation, equal to 2·09 volts in the case of sodium. The electron within the atom being now in an unstable orbit falls to the 1·5 S ring, and in so doing gives up the quantum of energy received during its ejection from the 1·5 S ring, as a quantum of radiation $h\nu$, with the emission of the D lines. When energy equal to or greater than 5·12 volts is communicated to the atom, the valency electron in the atom is displaced from the 1·5 S ring to infinity, and in returning may occupy any one of the multitude of stationary states, the dimensions of which are large compared with the orbits of the other electrons, within the atom, before it finally reaches the stable orbit. The many electrons returning in the different atoms will do so by various paths, and we obtain the many line spectra, the intensities of the lines representing to some approximation the chances of the inter-orbital transitions. As at present only two potentials, viz. the resonance potential and the ionizing potential, have been observed for sodium vapour, it seems probable that there are only two ways in which the electron may move within the atom when being ejected from the 1·5 S ring. It is ejected to the first p ring, or it is removed to infinity. This applies when electrical stimulus is employed.

The lines of the principal series of sodium are all absorption lines, and no line of any other series has been reversed, although recently Datta * has shown that the two combination lines 1·5 S $-$ $2d$ and 1·5 S $-$ $3d$ of potassium are absorbed when light from an electric arc is passed through potassium vapour. The excitation of these lines is well known, and previously it was thought that they appeared owing to the presence of the electrostatic field, as they represent inter-orbital transfers requiring two units change in azimuthal quantum, which is inadmissible according to the selection principle of Bohr. Foote, Meggers, and Mohler † have shown also that the combination line 1·5 S $-$ $2d$ may be

* Proc. Roy. Soc. A. vol. ci. (1922).

† Phil. Mag. vol. xliii. (1922).

produced in a low voltage arc where the external electric field cannot be effective. It has been pointed out by Bohr *, however, that this line was only excited in the experiments when the electron density was very great. In this case the fields due to the neighbouring positive ions and electrons, to which the emitting atoms are subject, may have been many times greater than the intensity of the external electric force present due directly to the applied potential difference.

The radiation quanta absorbed are able to remove the electron to any of the p orbits within the atom, and so the energy absorbed may be very nearly that required to ionize the atom. On the return of the electron to its stable orbit radiation energy is emitted, and is observed as fluorescence. In this way Rayleigh † has excited the fluorescence of sodium vapour by the 5896, 5890 lines, and also by the 3303 line. It thus appears possible to do by absorption what cannot be performed by electron impacts, namely, displace the electron to any of the p orbits.

Low voltage arcs in metallic vapours suggest that it is possible to eject the electron to infinity from the $2p$ ring, for these arcs operate at a potential below the ionizing potential. The energy necessary to eject the electron from this orbit will certainly be less than that required if the electron is in the 1·5 orbit. The energy necessary for the transfer to the $2p$ ring may be obtained from the bombardment by an electron having energy equal to the resonance potential, or by the absorption of radiation $h\nu$ from adjacent atoms which are in the partially ionized condition. Atoms of metallic vapours seem to permit of the radiant energy liberated by each electron impact which results in radiation, to be passed on from atom to atom, and this multiplies the fraction of atoms which are in the abnormal or partially ionized condition. The greater part of the radiation emitted when electrons fall from the $2p$ to the 1·5 S rings will escape from the tube, but some of it may be taken up by neighbouring atoms. This absorption effect will increase with the vapour-pressure, and with an increase in the density of the bombarding electrons. Under these conditions many of the valency electrons of the atoms will be in the $2p$ rings, and so absorption of the principal lines will not occur, but rather those lines belonging to series converging to $2p$ will become absorption lines. Theoretically then, if sodium vapour at a fairly high pressure is excited by a dense stream of electrons, and white light passed through the luminescent

* Phil. Mag. vol. xliii. (1922).

† Strutt, Proc. Roy. Soc. A. vol. xcvi. (1919).

vapour, the subordinate series lines should show absorption. This was first suggested by Foote and Mohler *.

Low voltage arcs always require a higher applied potential difference to start them than is necessary for their maintenance once they are started. After the initial discharge the ions present, together with the electrons, may produce internal electric fields far greater than the external applied electric force. On the theory of the radiation absorption, part of the energy for ionization is obtained from the radiation emitted by those atoms already ionized, so that less applied electric force is required after the initial discharge.

Pflüger †, using a condensed electric discharge in a three-electrode tube in which a short constriction provided the source of radiation and the wider and longer part the absorbing column, investigated the absorption and reversal of the hydrogen lines. He succeeded in reversing Hα. Kuch and Retschinsky ‡, Pflüger §, and Grebe || found that luminescent mercury vapour absorbed lines which were different from those absorbed by ordinary mercury vapour.

Robertson ¶, using a sodium vapour electric discharge-tube described by the author, excited fluorescence in sodium vapour contained in another quartz tube. White light was passed through the latter, and the absorption spectra noted when the sodium vapour was fluorescing, and when it was not. There was no change in the absorption spectra, and no absorption of the subordinate series lines occurred. This was probably due to the small amount of fluorescence produced, with the result that comparatively few of the atoms had the valency electron ejected to the $2p$ ring, and so few were in a condition to absorb radiation corresponding to those lines which converge to this limit.

2. *Experiments.*

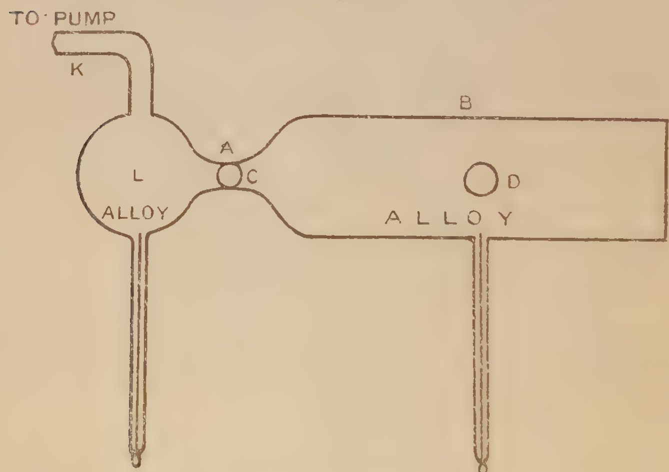
It was thought that an experiment similar to that performed by Pflüger with hydrogen, but using sodium vapour instead of the gas, might show peculiar absorption effects. A quartz tube about 50 cm. long with a narrow constriction at one end was used, and a condensed electric discharge was sent through the sodium vapour maintained at 300° C. The light from the constriction was viewed through a quartz

* Phil. Mag. vol. xl. (1920).
† *Ann. der Phys.* Bd. xxiv. (1907).
‡ *Ann. der Phys.* Bd. xxii. (1907).
§ *Ann. der Phys.* Bd. xxvi. (1908).
|| *Ann. der Phys.* Bd. xxxvi. (1911).
¶ 'Nature,' Jan. 1922, p. 43.

window at the end of the wider portion of the tube. The D lines alone showed self-reversal, and passing white light along the length of the tube simply gave absorption of the same lines. Here again the density of the free electrons within the tube is comparatively small. Another experiment using the principle described by the author * in the sodium-potassium vapour arc-lamp was tried.

The apparatus is shown in fig. 1. It was made of quartz. The constriction A was 3 mm. in diameter, and the wider portion B 2·5 cm. in diameter and 20 cm. long. The

Fig. 1.



end of the wide portion was closed by a quartz window E fused to the end of the tube. Side tubes at C and D closed by quartz windows permitted the observation of the arc at different places. The alloy of sodium and potassium was introduced into the apparatus at K, and after evacuation of the apparatus, the alloy was run from L into B. Warming the alloy in L caused the oxide formed on the surface to disappear, and in this way the two pools of alloy had clean surfaces. With this arrangement an arc could be struck between the two alloy electrodes with an applied potential difference as low as 30 volts. The arc was struck by warming the alloy, connecting one electrode to a small induction coil, and passing a momentary electric discharge. Sometimes the arc could be started by simply tilting the apparatus, and so making momentary contact between the

* Phil. Mag. vol. xliv. (1922).

two alloy electrodes. Once started the arc could be maintained for an indefinite period. The apparatus was placed inside an electric heater, so that the vapour pressure of the sodium within could be maintained at any desired value. Visual observations of the series produced were made at A, D, and E, and the spectrograms photographed with a constant deviation type of spectrometer, using Wratten panchromatic plates.

3. *Experimental Results.*

Under all experimental conditions the potassium lines were faint compared with those of sodium, but became relatively brighter as the applied potential difference was decreased. The radiations appearing at A had the characteristic colour of the D-line radiation, but observed at D the colour changed to a pale greenish-yellow. At A the D lines were very intense compared with all other lines, but at D the subordinate series lines of sodium and potassium appeared much brighter, although the concentration of the luminous sources must be much greater at A. This was most marked with high current densities, and is illustrated in the spectrograms shown in Pl. I. The relative intensity of the various lines when the radiation was observed through the window E was intermediate between that at A and D. As the current density decreased the relative brightness of the D lines increased. Raising the temperature of the apparatus to 450° C. caused strong self-reversal of the D lines as seen at E, but no self-reversal of any other lines. An electromagnet was arranged so that the luminous column in the constriction A was squeezed to that part of the tube away from the window C. This caused self-reversal of the D lines owing to the light passing through the intervening layer of cooler vapour before it emerged at the window C. Even with this arrangement, however, the D lines were still very bright compared with the other lines.

The radiation issuing at C had passed through a shorter length of absorbing vapour than that coming out at D and E, but the subordinate lines were very faint. It must be remembered, however, that the D light at E had suffered considerable absorption, since it had been transmitted through a considerable thickness of sodium vapour. It was to be expected, therefore, that the D lines would appear faint compared with the other lines.

The luminous sources in the constriction A are concentrated, and so the subordinate series lines should appear

intrinsically brighter at A than at D. Such is not the case. This can be explained if these lines are self-reversed, as seen at A. At the constriction the current density is high, and so the electrons are very dense. This is the condition for the absorption of the D-line radiation by atoms from neighbouring atoms. Many of the atoms present will thus be in the condition that the valency electron has been displaced to the $2p$ ring, and so these atoms are no longer in a condition to absorb the D-line radiation, but rather those lines converging to $2p$. Thus the light transmitted by these atoms suffers absorption of the subordinate lines, while the D lines are not affected. In the wider portion of the apparatus the electrons are less dense. There is less chance of the atoms absorbing radiation, and so the subordinate lines no longer show self-reversal. Accordingly, they will appear brighter than at the constriction.

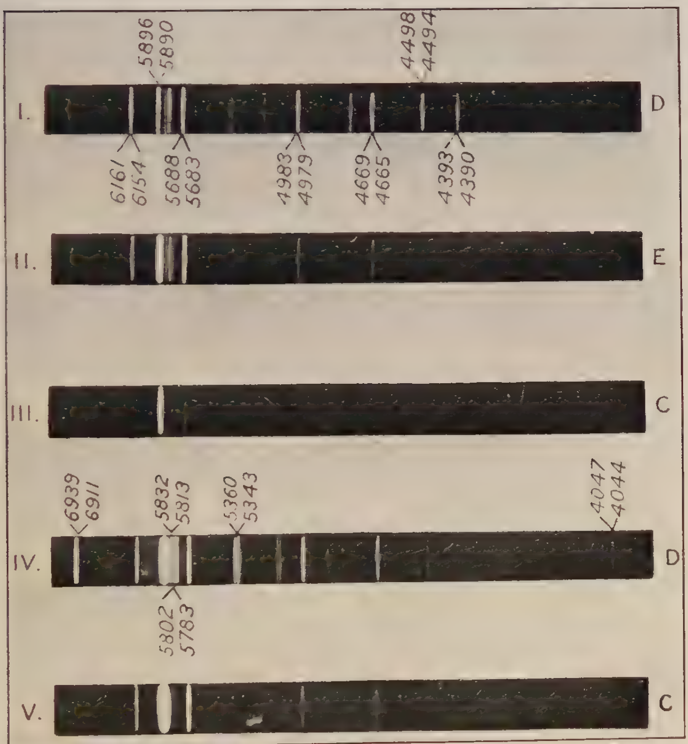
When the current in the arc is reduced, the density of the electrons and the self-reversal of the subordinate lines are less marked. This is illustrated in the spectrograms IV., V. (Pl. I.).

While the arc is in operation the actual drop of potential across the electrodes is only about 10 volts. At 300° C. the vapour pressure of sodium and potassium is high, and the mean free path of the electrons must be very small. It seems improbable that the bombarding electrons can gain energy sufficient to ionize during the time that this path is traversed. It is more likely that the atoms become partially ionized by the absorption of radiation. The atoms which are in this condition, *i. e.* those in which the valency electron is in the $2p$ ring, will then require less energy than 5·1 volts for complete ionization.

White light from an arc was passed through the radiation at the constriction, and also at D, and the complete spectrum viewed at the windows C and D. Visual observations showed absorption of the D lines only, they of course appearing dark on the luminous background.

The light issuing at D came from a point within the tube at a considerable height above the alloy electrode. There was no possibility of the variation in the relative intensity of the lines being due to the electrode potential gradient.

The spectrograms I., II., III. (Pl. I.) were photographed with exposures of 10 seconds each, and those of IV., V. (Pl. I.) with 20 seconds' exposure.



I., II., III. Current 5·1 amps.
IV., V. Current 2·0 amps.
D, E, C refer to fig. 1 and indicate where the spectrograms were photographed.

III. *Diffraction Pattern in a case of two very close Point-Light Sources.* *By* B. E. MOURASHKINSKY, *Optical Laboratory of the Central Chamber of Weights and Measures, Petrograd* *.

[Plate II.]

AS is well known, the illumination at the point r in the focal plane of a geometrically corrected object-glass with a circular aperture due to a point-light source is expressed by

$$I = M\frac{4J_1^2(z)}{z^2}, \quad \ldots \ldots \quad (1)$$

where

$$M = \frac{\pi^2 R^4}{\lambda^2 f^2}; \quad z = \frac{2\pi R}{\lambda f} r, \quad \ldots \ldots \quad (2)$$

R being the radius of the object-glass, f its focal length, λ the wave-length of light, r the distance of the considered point from the geometrical image of the point-source (in the focal plane), J_1 Bessel's function of order unity. The illumination at the geometrical image of the source is assumed to be equal to unity.

For values of

$$\frac{4J_1^2(z)}{z^2} \quad \ldots \ldots \quad (1\,a)$$

we have Lommel's tables, in which the argument is varying by tenths from $z=0$ to $z=20{\cdot}0$.

If we take two point-sources, primarily of equal intensity, and denote the illumination at the point r due to the first source by I_1 and that due to the second by I_2, the entire illumination at this point due to both the sources will be

$$I = I_1 + I_2,$$

$$I = M\left[\frac{4J_1^2(z_1)}{z_1^2} + \frac{4J_1^2(z)}{z^2}\right], \quad \ldots \ldots \quad (3)$$

$$z = \frac{2\pi R}{\lambda f} r,$$

$$z_1 = \frac{2\pi R}{\lambda f} r_1,$$

where r and r_1 are the distances of this point in the focal plane from the geometrical images of two sources.

* Communicated by the Author, having been read before the Russian Astronomical Society, April 27, 1922.

The distribution of illumination along a meridional line (a line joining two geometrical images) depends on the distance between the sources. If the distance between two geometrical images is D,

$$D = \frac{2\pi R}{\lambda f} d, \quad \ldots \ldots \ldots \quad (4)$$

d being a linear distance between these images in focal plane; then

$$r_1 = d - r,$$

$$z_1 = D - z.$$

The value z may be either positive (towards the image of the second point) or negative (in the opposite direction). The expression (3) may be written

$$I = M\left[\frac{4J_1^2(z)}{z^2} + \frac{4J_1^2(D-z)}{(D-z)^2}\right]. \quad \ldots \quad (5)$$

The factor M is constant for a given instrument, and the illumination at the point r will be

$$\frac{4J_1^2(z)}{z^2} + \frac{4J_1^2(D-z)}{(D-z)^2}. \quad \ldots \ldots \quad (6)$$

The measured angular distance between two point-sources expressed in arc seconds may be converted into D and *vice versa* by means of the following simple relations:

$$D = \frac{2\pi R}{\lambda}\frac{d}{f},$$

$$D = \frac{2\pi R}{\lambda}\frac{D''}{\tan 1''}, \quad \ldots \ldots \ldots \quad (7)$$

$$D'' = D \cdot \frac{\lambda}{2\pi R} \tan 1'' \quad \ldots \ldots \quad (8)$$

As two points are of equal intensity, the distribution of illumination along a meridional line will be symmetrical with respect to two geometrical images.

The question is, what value should D attain so that the two images would be just resolved.

The resolving power of an object-glass with a circular aperture depends on: (1) the size of an aperture, *i.e.* its diameter, (2) the wave-length of light of a source, (3) the contrast between the illumination at the geometrical images of two sources and that at the central minimum in the

diffraction-pattern. An object-glass with the radius R for one angular distance between two point-sources D″ gives the distance $D = \frac{2\pi R}{\lambda} \frac{D''}{\tan 1''}$; the other object-glass for the same angular distance gives another value of D. For each D we have the corresponding ratio of illumination (in the diffraction pattern) at its maximum to that at the central minimum. The eye can detect only a limit difference of illumination depending on its contrast sensibility.

The general theories of the resolving power state that two point-light sources of equal intensity are resolved when the maximum of illumination in the diffraction pattern due to one of them coincides with the first minimum in the diffraction pattern due to the other, *i. e.* when the distance between two geometrical images of such sources is equal to the radius of the first dark ring. Lord Rayleigh, in his paper "Wave Theory," admits as a limit of resolution a coincidence of the maximum of illumination in the diffraction pattern due to one object with the first minimum due to the other in all the cases of point and line sources for circular and rectangular apertures. In the case of two points we have the usual relations:

$$D = \frac{2\pi R}{\lambda} \frac{d}{f} = 3{\cdot}8316,$$

$$\tan D'' = \frac{d}{f} = \frac{3{\cdot}8316}{\pi} \cdot \frac{\lambda}{2R},$$

$$\tan D'' = 1{\cdot}22 \frac{\lambda}{2R}. \qquad \text{(9)}$$

This general relation is entered without any further consideration into many optical text-books, even into the best ones.

The resolving power of an object-glass is generally considered as the least angular distance between two objects at which the latter are just resolved, easily resolved, etc. These last terms introduce a great uncertainty as to what exactly is meant by resolution. If we admit the resolution as a complete separation (contact) of the central bright spots in the diffraction pattern, we must know the diameter of the visible part of the central spots, because we do not observe a first dark ring as a geometrical circumference, but it has a sensible width. The diameter of the visible part of a central spot depends on the threshold sensibility of the eye for a case of the diffraction pattern; and this depends on

the brightness of a source, on its colour (on the wave-length of light in a case of a monochromatic source), and on the brightness and colour of the background. Considering the resolving power as a contact of the visible parts of two central spots, we see that the relation (9) gives that the part of the central spot perceptible to the eye is only about half a diameter of the first dark ring. Even if this be correct, it will be so only for quite a determined value of intensity and wave-length of light of a source, and cannot be applied *a priori* to all sources whatever their brighiness and colour.

In a case of two point-sources, we can often consider two points as resolved though their central spots partly overlap each other.

If we assume the resolution as the least distance between two sources at which a distribution of illumination in the diffraction pattern due to both the sources is such that the difference between illumination at its maximum (on the meridional line) and that at central minimum is just perceptible to the eye, the expression (9) does not represent the conditions of resolution, as in this case the percentage difference of illumination is equal 26·5 (see below) and is much inferior to the limit of the contrast sensibility of the eye.

Different authors give different values for the contrast sensibility of the eye in a case of diffraction pattern. Thus Strehl * gives 3–4 per cent., and found for this difference the least distance between two points just resolved to be D=3·2.

Wadsworth † in one of his papers on resolving power says:—"The first (resolving power) defines, as is well known, the power of an objective to 'resolve' or show as separate objects two or more close lines or points . . . , and the limit of resolution (angle between two points or fine lines that just appear separated) is represented by the well-known expression

$$a = m\frac{\lambda}{b},$$

where m is variable from unity for a rectangular aperture (of width b) to about 1·1 for a circular aperture of diameter b." We have for this case D=3·14 for a rectangular aperture and D=3·55 for a circular aperture. Wadsworth's

* Strehl, *Zeit. für Instr.* xvi. p. 259 (1896). No details of this assumption are found in Strehl's paper.

† Wadsworth, 'Observatory,' xx. p. 333 etc.

data are taken from Lord Rayleigh's "Wave Theory," but they relate only to line sources which, in fact, have a first minimum at 3·14 and 3·55 for the rectangular and circular aperture respectively. Classen * assumes the contrast sensibility of the eye in the case of diffraction pattern to be equal to 5 per cent., without any indication of the basis of this assumption. For this difference the distance between two point-sources will be $D=3·3$.

The contrast sensibility of the eye has in many cases been investigated in laboratory conditions: these investigations give the concordant results on the sensibility of the eye for different conditions of brightness. From König's results, recalculated by Branshard †, we see that the least difference of the field brightness just perceptible to the eye in the case of two adjacent fields is 0·017 for the field brightness equal to about 200 millilamberts. For a field brightness greater or less than this, the least difference just perceptible to the eye is greater than 0·017.

In many cases observers of double stars are able to resolve (even to measure a distance between) two stars, the distance between which expressed as D is much less than $D=3·83$, even than $D=3·3$.

We give here some data from several catalogues of double stars observed with different instruments. These data represent (for the given instrument) the least measured or estimated distances between two double stars' components of equal or nearly equal brightness. The distances obtained by measuring the two diameters of a complex oblong figure without resolving two components are not considered here.

Of course (7) cannot be strictly applied to observations of stars, as it refers to a monochromatic source of light and to a geometrically corrected object-glass. A determination of the resolving-power of an objective by astronomical observations is greatly influenced by atmospherical conditions, general absorption of atmosphere, brightness of a background; besides that the least distances at which two stars are resolved are taken from the measurement of these distances made with the same instrument, the resolving-power of which is to be found. For these reasons, the following values of D may be regarded as approximate ones. These values of D are calculated by means of (7), λ being taken equal to 0·00055 mm.

* Classen, *Mathematische Optik*, p. 201.
† Branshard, Phys. Review, Feb. 1918.

The values of the distances put in brackets are estimated values. In the square brackets are given the stellar magnitudes of the components. The data marked D_c, %, and D_c'' are to be explained later.

		D″.	D.	D_c.	%.	D_c''.
1. Burnham (36 inch). (Publ. of Lick Obsy. ii. p. 103.) Σ 2367 [7·2–7·6].	1891 326	0″·07 ±	1·8	3·11	1·0	0″·12
	389	·09	2·25	3·18	2·5	·13
	578	·09	2·25	3·18	2·5	·13
	1891 43	0″·09	2·25	3·18	2·5	0″·13
2. Comstock (15·5 inch). (Publ. of the Washburn Obsy. x. pt. 1, p. 48.) β 814 [8·5–8·5 (β)].	1893 570	0″·21	2·2	3·17	2·5	0″·29
	95 628	·23	2·4	3·21	3·0	·29
	96 430	·21	2·2	3·17	2·5	·29
3. Comstock (15·5 inch). (Ibid. p. 59.) Σ 2315 [7–8·0 OΣ Aeq–HΣ].	1895 805	0″·21	2·2	3·17	2·5	0″·29
	96 665	·21	2·2	3·17	2·5	·29
4. O. Struve (15 inch). (Obs. de Poulkova, ix. Suppl. p. 16.) Σ 1728 [6·0–6·0]. 42 Comæ Ber.	1847 45	(0″·22)	2·3	3·19	3·0	0″·30
	48 38	·25	2·6	3·27	5·0	·31
	48 46	·26	2·7	3·30	6·0	·31
5. O. Struve (15 inch). (Ibid. p. 39.) OΣ 208 [5–5·6]. φ Urs. Maj.	1872 41	(0″·23)	2·4	3·21	3·5	0″·30
	72 42	(·25)	2·6	3·27	5·0	·31
6. Burnham (12 inch). (Publ. of Lick Obsy. ii. p. 124.) OΣ 535 [7–7·5 (OΣ)].	1888 652	0″·24	2·0	3·14	2·0	0″·37
	655	·23	1·9	3·12	1·5	·37
	714	·29	2·4	3·22	3·0	·38
	733	...	...	...	...	...
	1888 69	0″·25	2·1	3·15	2·0	0″·37
7. W. Struve (4·9 inch). (Mensuræ Micr. p. 11.) Σ 278 [8·4–8·7].	1827 26	0″·52	3·5	3·60	...	0″·53
	31 27	(·35)	2·4	3·21	3·0	·47
	31 32	(·35)	2·4	3·21	3·0	·47
	33 23	(·50)	3·4	3·56	...	·53
	1830 77	0″·430	2·9	3·36	8·0	0″·50

To investigate the question more closely, the values of I are calculated from (6) for the series of points varying in distance from the geometrical image of each point-source by tenths from $-0·5$ to $D+0·5$ for the different distances between the point-sources from $D=2·9$ to $D=5·1$. The tables of all the values of I are rather long to be given here

in full; the values for the most interesting points (at the geometrical images, at the maxima and the central minima) are given in Table I. The values of I for some distances are plotted in Pl. II. fig. 1, where M_1 and M_2 represent the positions of maxima, O_1 and O_2 the positions of the geometrical images.

TABLE I.

D.	I_0.	I max.	I min.	$\frac{\text{I min.}}{\text{I max.}} . 100.$
2·9	1·0670	1·1528	...	...
3·0	1·0511	1·1073	1·1068	99·95
3·1	1·0377	1·0715	1·0607	99·0
3·2	1·0267	1·0463	1·0149	97·0
3·3	1·0179	1·0288	0·9692	94·2
3·4	1·0111	1·0167	0·9240	91·4
3·5	1·0062	1·0086	0·8794	87·2
3·6	1·0028	1·0037	0·8350	83·2
3·7	1·0008	1·0008	0·7912	79·0
3·8	1·0000	1·0000	0·7350	73·5
3·9	1·0002		0·7063	70·6
4·0	1·0011		0·6652	66·5
4·1	1·0025		0·6250	62·3
4·2	1·0044		0·5859	58·3
4·3	1·0064		0·5478	54·4
4·4	1·0085		0·5109	50·7
4·5	1·0105		0·4752	47·0
4·6	1·0124		0·4408	43·5
4·7	1·0141		0·4077	40·2
4·8	1·0155		0·3758	37·0
4·9	1·0165		0·3454	34·0
5·0	1·0172		0·3163	31·1
5·1	1·0175		0·2886	28·4

From the data of Table I. we distinguish three cases, namely:

(1) The distances between two point-sources $D > 3·83$.
(2) The distances from $D = 3·83$ to $D = 3·0$.
(3) The distances $D < 3·0$.

In the first case we have two maxima of illumination in the diffraction pattern. The positions of both the maxima correspond to the two geometrical images of the two sources, and vary in values depending on a distance between the sources according to variations of the value of $1·0000 + I(D)$:

For $D = 3·83$ we have $I = 1·0000$;
,, $D = 5·1$,, ,, $I = 1·0175$; etc.

The least difference between the illumination at maxima (at the geometrical images) and that at the point between them is 26·5 per cent. for the distance D=3·83. This difference is greater than necessary to be just perceptible to the eye. Two point-sources in this case are easily resolved as the general theory requires.

In the second case we have *two maxima also, but their positions do not correspond to the positions of the geometrical images.* For the distance D=3·83 the displacement of the maxima does not as yet occur. As D decreases, the displacements of the maxima become more and more sensible. Both the maxima are displaced symmetrically towards the centre of the diffraction pattern.

The positions of the maxima for each given distance D may be determined as the roots of the equation

$$\frac{4J_1^2(z)}{z^2}+\frac{4J_1^2(D-z)}{(D-z)^2}=\max. \quad . \quad . \quad . \quad . \qquad (10)$$

The displacements of the maxima for the different distances D are given in Table II. and Pl. II. fig. 2.

Table II.

D.	Δz.	$2\Delta z$.
3·8	0	0
3·7	>0	>0
3·6	0·05	0·1
3·5	0·1	0·2
3·4	0·2	0·4
3·3	0·3	0·6
3·2	0·4	0·8
3·1	0·7	1·4
3·0	1·1	2·2

We see that the displacements of the maxima of illumination increase very rapidly as the distance decreases.

In measuring (micrometrically) the distance between such very close point-sources we tend to set the cross wire certainly not at the geometrical images (which may be unknown) but at maxima of illumination which occur evidently on the meridional line in the diffraction pattern.

If the magnification of the instrument used would allow the measurements of the distance between two maxima in an image of two point-sources with such an accuracy that the errors of the measured distance would be much less than the displacement of maxima themselves, we must correct the measured distance to obtain the real one.

If we observe two close stars of equal brightness, their measured distance will be less than the real one, and *vice versa*, if we found from the measurements the distance between two point-sources to be less than D=3·83 their real distance will be greater. To find the real distances we have to modify the Table II. so as to get the values of the real distances for the different measured distances.

TABLE III.

D.	D_c.	D.	D_c.	D.	D_c.
0·8	3·00	1·9	3·12	2·9	3·36
0·9	3·01	2·0	3·14	3·0	3·39
1·0	3·02	2·1	3·15	3·1	3·43
1·1	3·03	2·2	3·17	3·2	3·47
1·2	3·04	2·3	3·19	3·3	3·51
1·3	3·05	2·4	3·21	3·4	3·56
1·4	3·06	2·5	3·23	3·5	3·60
1·5	3·07	2·6	3·27	3·6	3·67
1·6	3·08	2·7	3·30	3·7	3·73
1·7	3·10	2·8	3·33	3·8	3·80
1·8	3·11				

If the necessary accuracy as just determined cannot be obtained, the errors of the measured distance may be of the same order as the values of the displacement of maxima (particularly for the real distances near D=3·8) and the corrections from the Tables II. and III. will not give the correct value of this distance. For the observed distance between the components of double stars given on p. 34 expressed in terms of D we find from the Table III. the corrected distances D_c and by (8) the corrected distances in arc seconds D_c''.

The values of the maxima of illumination in this case (3·83 > D > 3·0) are given in the first table in column 3. We see that for these distances the maxima of illumination vary in value, increasing (as D decreases) from unity for D=3·83 to about 1·1 for D=3·0.

The difference between the illumination at maxima (not at the geometrical images now) and that at the central point of the diffraction pattern decreases with the distance between the point-sources. The values of illumination at the centre of the diffraction pattern as referred to that at maxima assumed to be equal 100 are given in the fifth column of Table I. Pl. II. fig. 3 represents these variations of difference of illumination at the mentioned points.

This difference varies from 26·5 per cent. for D=3·83 to 0·05 per cent. for D=3·0. These differences for the distances near D=3·0 are hardly, if at all, perceptible to the eye.

In the third case (D < 3·0) we have only one maximum at the point midway between two geometrical images: no central minimum exists in this case. As the difference between the illumination at its maximum and that at its central minimum in the case of D=3·0 is only 0·05 per cent., this distance can be considered as the critical distance. In this case the difference between the illumination at its maximum and that at the geometrical images is 5 per cent. If we neglected the displacement of maxima and referred the illumination at the central point to the illumination at the geometrical images (instead of that at maximum) we should consider the distance D=3·16 (instead of D=3·0) as the critical distance, as for this case the difference of illumination is equal to about zero.

As D decreases, the maximum of illumination being always at the point midway between two geometrical images increases in value from 1·1073 for D=3·0 to 2·0000 for D=0 (*i.e.*, two geometrical images coincide).

Thus this critical distance D=3·0 must be considered as the absolute limit of resolution of two very close point-light sources of equal intensity.

In the table of the least measured distances between the components of double stars, we find under D the values less than D=3·3 and even less than D=3·0. We have just seen that the last distance must be considered as the absolute limit of resolution. Having introduced the values of the displacements of the maxima and written for each D the corresponding corrected values D_c, we shall find that these distances are not less than D=3·15 or 3·20. For these we have the difference between the illumination at its maximum and that at the central minimum equal to 2 or 3 per cent. respectively.

We have seen that the contrast sensibility of the eye in the most favourable conditions of brightness (200 ml.) does not exceed 1·7 per cent., and that only for the adjacent fields. In the case of a diffraction pattern we do not deal with two adjacent fields, and therefore the least difference of illumination just perceptible to the eye will differ from that limit value.

Therefore, if our assumption of the necessary accuracy of the measurements were fulfilled and thus the observed

distances could be converted by means of the data of Table III. into corrected ones, we might conclude from the observations of the double stars that the least difference between the illumination at its maximum and that at the central point in the diffraction pattern due to two very close point-sources of equal intensity, just perceptible to the eye, would be not greater than 2–2·5 per cent.

As we have said, the astronomical observations are greatly dependent on atmospherical conditions, and the real least difference of illumination for the case in question (diffraction pattern due to two very close point-sources) just perceptible to the eye may be less than that found from the observations of double stars.

In a case of two point-sources of unequal intensity, the illumination at a point in the diffraction pattern in the focal plane of an objective will be expressed by

$$I = \frac{4J_1^2(z)}{z^2} + n\frac{4J_1^2(D-z)}{(D-z)^2}, \quad . \quad . \quad . \quad . \quad (11)$$

where n is the ratio of intensities of two point-sources.

If we assume the intensity of a brighter point as unity, n will be less than unity and we have

$$I_2(z) = \frac{1}{n} I_1(z).$$

To represent the distribution of illumination along a meridional line, the values of I were calculated for the different D from $D=3·0$ to $D=5·1$, and for the different n from $n=0·9$ to $n=0·01$. These calculations were made as before, and all data are expressed in the illumination at the geometrical image of the brighter point being assumed to be equal to unity.

The whole table is too long to be given at length, and as before we give here but the final results (Table IV). In this table I_1, I_m, I_2 denote the illumination at a first maximum, that at minimum on a meridional line, and at a second maximum. In the last column are found the percentage ratios of illumination at its minimum to that at a second maximum. The positions of the minima (their distance from the geometrical image of a fainter point) are added in brackets in column I_m. The values of the displacements of the maxima, if the latter exist, are given also in brackets in columns I_1 and I_2.

Table IV.

	$n = 0·9.$				$n = 0·8.$			
D.	I_1.	I_m.	I_2.		I_1	I_m.	I_2.	
3·2	1·0380(+0·35)	0·9529(1·0)	0·9533(+0·7)	99·92				
3·3	1·0239(+0·2)	0·9131(1·3)	0·9316(+0·4)	98·0				
3·4	1·0140(+0·2)	0·8721(1·4)	0·9177(+0·25)	95·0	1·0118(+0·1)	0·8068(1·1)	0·8198(+0·3)	98·41
3·5	1·0075(+0·1)	0·8309(1·6)	0·9089(+0·2)	91·4	1·0064(+0·1)	0·7714(1·3)	0·8099(+0·2)	95·2
3·6	1·0020(+0·1)	0·7893(1·6)	0·9039(+0·1)	87·3	1·0025(+0·1)	0·7350(1·4)	0·8042(+0·1)	91·4
3·7	1·0008	0·7484(1·7)	0·9008	83·1	1·0006	0·6982(1·5)	0·8009(+0·05)	87·2
3·8	1·0000	0·7083(1·8)	0·9000	78·7	1·0000	0·6614(1·6)	0·8001	82·7
3·9	1·0002	0·6694(1·9)	0·9002	74·4	1·0002	0·6250(1·7)	0·8002	78·1
4·0	1·0010	0·6297(1·9)	0·9011	69·9	1·0009	0·5894(1·8)	0·8011	73·6
4·1	1·0023	0·5922(1·95)	0·9025	65·6	1·0020	0·5544(1·85)	0·8026	69·1
4·2	1·0039	0·5547(2·0)	0·9044	61·3	1·0043	0·5198(1·9)	0·8044(−0·05)	64·6
4·3	1·0058	0·5190(2·1)	0·9064	57·3	1·0051	0·4867(2·0)	0·8065(−0·1)	60·3
4·4	1·0076	0·4840(2·1)	0·9085	53·3	1·0068	0·4540(2·0)	0·8086(−0·1)	56·1
4·5	1·0095	0·4503(2·2)	0·9105	49·5	1·0084	0·4222(2·1)	0·8186	52·1
4·6	1·0112	0·4178(2·2)	0·9124	45·8	1·0099	0·3922(2·2)	0·8124	48·3
4·7	1·0127	0·3862(2·3)	0·9141	42·2	1·0112	0·3625(2·2)	0·8141	44·5
4·8	1·0139	0·3564(2·3)	0·9155	38·9	1·0124	0·3345(2·3)	0·8154	41·0
4·9	1·0148	0·3272(2·4)	0·9165	35·7	1·0132	0·3075(2·3)	0·8165	37·7
5·0	1·0154	0·3003(2·4)	0·9172	32·7	1·0138	0·2815(2·4)	0·8172	34·4
5·1	1·0157	0·2735(2·5)	0·9175	29·8	1·0140	0·2574(2·4)	0·8174	31·5

	$n = 0{\cdot}7.$				$n = 0{\cdot}6.$			
3·5	1·0053(+0·1)	0·7003(1·0)	0·7111(+0·25)	98·48	1·0017	0·5953(0·9)	0·6051(+0·2)	98·38
3·6	1·0020	0·6707(1·2)	0·7015(+0·1)	95·2	1·0005	0·5710(1·1)	0·6013(+0·1)	95·0
3·7	1·0006	0·6394(1·3)	0·7011(+0·1)	91·2	1·0000	0 5450(1·2)	0·6000	90·8
3·8	1·0000	0·6075(1·4)	0·7000	86·8	1·0001	0 5177(1·4)	0·6002	86·2
3·9	1·0001	0·5753(1·5)	0·7002	82 2	1·0007	0·4901(1·5)	0·6011(−0·05)	81·5
4·0	1·0008	0·5431(1·6)	0·7011	77·5	1·0015	0·4626(1·6)	0 6029(−0·1)	76·7
4·1	1·0015	0·5113(1·7)	0·7027(−0·1)	72·8	1·0026	0·4353(1·7)	0·6049(−0·1)	72·0
4·2	1·0030	0·4801(1·8)	0·7047(−0·1)	68·1	1·0038	0·4084(1·7)	0·6070(−0·1)	67·3
4·3	1·0045	0·4499(1·9)	0·7068(−0·1)	63·7	1·0051	0·3817(1·8)	0·6090(−0·1)	62·7
4·4	1·0059	0·4201(1·9)	0·7088(−0·1)	59·3	1·0063	0·3557(1·9)	0·6109(−0·1)	58·2
4·5	1·0074	0·3909(2·0)	0·7107(−0·1)	55·0	1·0074	0·3308(1·95)	0·6126(−0·05)	54·0
4·6	1·0087	0·3632(2·1)	0·7124(−0·05)	51·0	1·0085	0·3066(2·0)	0·6141	49·9
4·7	1·0099	0·3363(2 1)	0·7141	47·1	1·0093	0·2827(2·1)	0·6155	45·9
4·8	1·0108	0·3100(2·2)	0 7155	43·3	1·0099	0·2602(2·2)	0·6165	42·2
4·9	1·0115	0·2855(2·3)	0·7165	39·8	1·0103	0·2588(2·2)	0·6172	38·7
5·0	1·0120	0·2613(2·3)	0·7172	36·4	1·0105	0·2178(2·3)	0 6175	35·3
5·1	1·0122	0·2399(2·3)	0·7175	33·4				

	$n = 0{\cdot}5.$				$n = 0{\cdot}4.$			
3·7	1·0004	0·4919(0·7)	0·5016(+0·1)	98·07	1·0003	0·4022(0·4)	0·4022(+0·1)	100·0
3·8	1·0000	0·4732(1·0)	0·5000	91·6	1·0000	0·3911(0·7)	0·4000	97·8
3·9	1·0001	0·4519(1·2)	0·5002	90·3	1·0001	0·3766(0·9)	0 4002	94·1
4·0	1·0005	0·4295(1·3)	0·5013(−0·1)	85·7	1·0004	0·3603(1·1)	0·4015(−0·1)	89·7
4·1	1·0013	0·4067(1·4)	0·5032(−0·1)	80·8	1·0010	0·3429(1·2)	0·4034(−0·1)	85·0
4·2	1·0022	0·3837(1·5)	0·5052(−0·1)	76·0	1·0017	0·3250(1·3)	0 4054(−0·1)	80·2
4·3	1·0032	0·3607(1·6)	0·5073(−0·1)	71·1	1·0026	0·3068(1·4)	0·4079(−0·1)	75·2
4·4	1·0042	0·3380(1·75)	0 5093(−0·1)	66·4	1·0034	0·2884(1·5)	0·4096(−0·1)	70·4
4·5	1·0053	0·3137(1·8)	0·5112(−0·1)	61·4	1·0042	0·2700(1·6)	0·4114(−0·1)	65·6
4·6	1·0062	0·2923(1·8)	0·5129(−0·1)	57·0	1·0050	0·2518(1·7)	0·4131(−0·1)	61·0
4·7	1·0071	0·2727(1·9)	0·5143(−0·1)	53·0	1·0056	0·2340(1·8)	0·4145(−0·1)	56·5
4·8	1·0077	0·2519(2·0)	0·5155(−0·05)	48·9	1·0062	0·2167(1·9)	0·4155(−0·05)	52·2
4·9	1·0082	0·2321(2·1)	0·5165	44·9	1·0066	0·2001(2·0)	0·4164	48·1
5·0	1·0086	0·2133(2·2)	0·5172	41·2	1·0069	0·1841(2·0)	0·4172	44·1
5·1	1·0087	0·1947(2·2)	0·5175	37·6	1·0070	0·1704(2·1)	0·4175	40·8

TABLE IV. (*continued*).

	$n = 0·3.$				$n = 0·2.$			
D.	I_1.	I_m.	I_2.		I_1.	I_m.	I_2.	
3·8	1·0000	0·2993(0·4)	0·3001(+0·1)	99·73				
3·9	1·0001	0·2916(0·7)	0·3003(−0·1)	97·1	1·0000	0·1983(0·4)	0·2006(−0·1)	98·85
4·0	1·0003	0·2818(0·8)	0·3018(−0·1)	93·4	1·0002	0·1938(0·6)	0·2024(−0·2)	95·8
4·1	1·0008	0·2700(1·0)	0·3036(−0·1)	88·9	1·0005	0·1875(0·7)	0·2044(−0·2)	91·7
4·2	1·0013	0·2576(1·1)	0·3056(−0·1)	84·3	1·0009	0·1806(0·9)	0·2065(−0·2)	87·05
4·3	1·0019	0·2446(1·25)	0·3077(−0·1)	79·5	1·0013	0·1728(1·0)	0·2086(−0·2)	82·8
4·4	1·0025	0·2309(1·4)	0·3098(−0·1)	74·5	1·0017	0·1646(1·1)	0·2104(−0·2)	78·2
4·5	1·0032	0·2171(1·5)	0·3117(−0·1)	69·7	1·0021	0·1557(1·3)	0·2121(−0·2)	73·4
4·6	1·0037	0·2033(1·6)	0·3134(−0·1)	64·8	1·0025	0·1466(1·4)	0·2136(−0·1)	68·6
4·7	1·0042	0·1897(1·7)	0·3147(−0·1)	60·3	1·0028	0·1374(1·5)	0·2150(−0·1)	63·9
4·8	1·0046	0·1763(1·75)	0·3157(−0·1)	55·8	1·0031	0·1282(1·6)	0·2160(−0·1)	59·4
4·9	1·0050	0·1629(1·8)	0·3165	51·5	1·0033	0·1191(1·7)	0·2167(−0·1)	55·0
5·0	1·0051	0·1500(1·9)	0·3172	47·3	1·0034	0·1103(1·8)	0·2172	50·8
5·1	1·0052	0·1375(2·0)	0·3175	43·3	1·0035	0·1015(1·9)	0·2175	46·7
	$n = 0·1.$				$n = 0·075.$			
3·7					1·0000	0·0744(−0·3)	0·0748(−0·55)	99·47
3·8	1·0000	0·0999(−0·1)	0·1004(−0·4)	99·52	1·0000	0·0750(0·05)	0·0770(−0·6)	97·4
3·9	1·0000	0·0998(0·1)	0·1003(−0·4)	97·4	1·0000	0·0749(0·1)	0·0790(−0·6)	94·8
4·0	1·0001	0·0986(0·3)	0·1046(−0·4)	94·3	1·0000	0·0742(0·3)	0·0810(−0·5)	91·6
4·1	1·0002	0·0969(0·4)	0·1066(−0·4)	91·0	1·0002	0·0729(0·4)	0·0829(−0·5)	87·9
4·2	1·0004	0·0941(0·6)	0·1085(−0·4)	86·7	1·0003	0·0713(0·55)	0·0845(−0·45)	84·4
4·3	1·0006	0·0912(0·75)	0·1102(−0·3)	82·8	1·0005	0·0691(0·7)	0·0862(−0·4)	80·2
4·4	1·0008	0·0875(0·9)	0·1119(−0·3)	78·2	1·0006	0·0666(0·8)	0·0875(−0·4)	76·1
4·5	1·0010	0·0836(1·0)	0·1132(−0·3)	73·9	1·0008	0·0638(0·9)	0·0888(−0·3)	71·8
4·6	1·0012	0·0795(1·1)	0·1144(−0·2)	69·5	1·0009	0·0609(1·0)	0·0898(−0·3)	67·8
4·7	1·0014	0·0752(1·2)	0·1155(−0·2)	65·1	1·0011	0·0578(1·1)	0·0908(−0·2)	63·7
4·8	1·0015	0·0706(1·3)	0·1162(−0·1)	60·1	1·0012	0·0545(1·25)	0·0914(−0·2)	59·6
4·9	1·0016	0·0661(1·4)	0·1169(−0·1)	56·5	1·0012	0·0512(1·4)	0·0920(−0·1)	55·7
5·0	1·0017	0·0615(1·5)	0·1172	52·5	1·0013	0·0477(1·5)	0·0923(−0·1)	51·7

	$n = 0{\cdot}05.$				$n = 0{\cdot}025.$			
D.	I_1.	I_m.	I_2.		I_1.	I_m.	I_2.	
3·2					1·0007	0·0223(−0·7)	0·0282(−1·6)	79·1
3·3					1·0004	0·0230(−0·6)	0·0293(−1·5)	78·5
3·4	1·0005	0·0467(−0·7)	0·0472(−1·05)	98·94	1·0003	0·0237(−0·5)	0·0305(−1·4)	77·7
3·5	1·0003	0·0480(−0·5)	0·0493(−1·0)	97·4	1·0002	0·0242(−0·4)	0·0316(−1·3)	76·6
3·6	1·0001	0·0491(−0·3)	0·0512(−0·9)	95·9	1·0001	0·0246(−0·3)	0·0327(−1·2)	75·2
3·7	1·0000	0·0499(0)	0·0531(−0·85)	94·0	1·0000	0·0249(−0·2)	0·0338(−1·1)	73·7
3·8	1·0000	0·0500(0)	0·0550(−0·8)	91·0	1·0000	0·0250(0)	0·0348(−1·0)	71·8
3·9	1·0000	0·0499(0·1)	0·0566(−0·75)	88·2	1·0000	0·0250(0·1)	0·0359(−1·0)	69·6
4·0	1·0001	0·0495(0·2)	0·0583(−0·7)	84·9	1·0000	0·0248(0·2)	0·0368(−0·9)	67·4
4·1	1·0001	0·0489(0·4)	0·0598(−0·6)	81·8	1·0001	0·0245(0·3)	0·0378(−0·9)	64·8
4·2	1·0002	0·0478(0·5)	0·0611(−0·6)	78·2	1·0001	0·0241(0·4)	0·0386(−0·7)	62·4
4·3	1·0003	0·0465(0·6)	0·0624(−0·5)	74·5	1·0002	0·0235(0·5)	0·0393(−0·6)	59·8
4·4	1·0004	0·0450(0·7)	0·0635(−0·4)	70·9	1·0002	0·0229(0·6)	0·0400(−0·6)	57·2
4·5	1·0005	0·0434(0·8)	0·0645(−0·4)	67·3	1·0003	0·0221(0·8)	0·0406(−0·5)	54·4
4·6	1·0006	0·0415(0·95)	0·0654(−0·3)	63·5	1·0003	0·0212(0·9)	0·0412(−0·4)	51·5
4·7	1·0007	0·0395(1·1)	0·0661(−0·3)	59·8	1·0004	0·0202(1·0)	0·0416(−0·3)	48·6
4·8	1·0008	0·0373(1·2)	0·0667(−0·2)	55·9	1·0004	0·0192(1·1)	0·0419(−0·25)	45·8
4·9	1·0008	0·0351(1·3)	0·0670(−0·1)	52·4	1·0004	0·0181(1·2)	0·0422(−0·2)	42·9
5·0	1·0009	0·0329(1·4)	0·0674	48·8	1·0004	0·0170(1·3)	0·0424(−0·1)	40·1
5·1	1·0009	0·0305(1·5)	0·0675	45·2	1·0004	0·0158(1·4)	0·0425	37·2

$n = 0{\cdot}01.$

D.	I_1.	I_m.	I_2.		D.	I_1.	I_m.	I_2.	
3·0	1·0005	0·0079(−1·1)	0·0205(−2·0)	38·5	4·1	1·0000	0·0098(0·2)	0·0253(−0·9)	38·7
3·1	1·0004	0·0087(−0·8)	0·0209(−1·9)	41·6	4·2	1·0000	0·0097(0·4)	0·0257(−0·8)	37·7
3·2	1·0003	0·0090(−0·7)	0·0213(−1·8)	42·3	4·3	1·0001	0·0094(0·5)	0·0260(−0·7)	36·2
3·3	1·0002	0·0093(−0·6)	0·0216(−1·7)	43·1	4·4	1·0001	0·0092(0·6)	0·0263(−0·65)	35·0
3·4	1·0001	0·0096(−0·5)	0·0222(−1·6)	43·2	4·5	1·0001	0·0089(0·7)	0·0266(−0·6)	33·5
3·5	1·0000	0·0098(−0·4)	0·0227(−1·5)	43·2	4·6	1·0001	0·0085(0·8)	0·0269(−0·5)	31·6
3·6	1·0000	0·0099(−0·2)	0·0232(−1·4)	42·7	4·7	1·0001	0·0082(0·9)	0·0271(−0·4)	30·3
3·7	1·0000	0·0100(−0·1)	0·0236(−1·3)	42·4	4·8	1·0002	0·0078(1·0)	0·0273(−0·3)	28·6
3·8	1·0000	0·0100(0)	0·0241(−1·2)	41·5	4·9	1·0002	0·0074(1·1)	0·0274(−0·2)	27·0
3·9	1·0000	0·0100(0)	0·0245(−1·1)	40·8	5·0	1·0002	0·0069(1·2)	0·0275(−0·1)	25·1
4·0	1·0000	0·0099(0·15)	0·0249(−1·0)	39·8	5·1	1·0002	0·0065(1·3)	0·0275	23·6

In the diffraction pattern due to two point-sources of unequal intensity we have two maxima of illumination different in magnitude corresponding to two sources, and a minimum between them which is not in the centre of the diffraction pattern, but nearer to the geometrical image of a fainter source as n decreases, and even may lie on the other side of the geometrical image of this source.

As in the previous case, the displacements of the maxima occur at very small distances between the sources, but in the case of two sources of unequal intensity the displacements of the first (greater) maximum are less than those of the second (less) maximum. The influence of light of the fainter point on the position of maximum of illumination in the diffraction image of the brighter point is greater in the case of smaller differences of intensity of two point-sources; at $n=0{\cdot}6$ and less the fainter point does not influence the position of the first maximum, and for these differences of intensity ($n<0{\cdot}7$) the first maximum is always at the geometrical image of the brighter point. The greatest displacement of the first maximum is $+0{\cdot}35$ for $n=0{\cdot}9$ and $\mathrm{D}=3{\cdot}2$.

As to the second maximum, its displacements resemble those for $n=1{\cdot}0$, but in absolute value they are always less than those for the same distance. As n decreases the displacements of the second maximum decrease also and after reaching the value zero take the opposite sign (negative displacement), increase in value, and then again decrease to zero. As n becomes less the negative displacements of the second maximum occur for a greater number of values of D. For $n=0{\cdot}2$ and less the displacements of the second maximum are only negative.

In order to resolve two points of unequal intensity, the eye has to detect the difference in the illumination at its minimum on a meridional line and that at the second (less) maximum. The percentage ratios of the illumination at these points are given in the last column of Table IV.

The critical distances, as might be expected, are greater with decreasing n; they are given in Table V.

TABLE V.

n.	D.	n.	D.
1·0	3·00	0·3	3·78
0·9	3·20	0·2	3·87
0·8	3·34	0·1	3·76
0·7	3·45	0·075	3·66
0·6	3·54	0·05	3·34
0·5	3·63	0·025	—
·	3·70	0·01	—

These values of D are obtained by the graphical extrapolation from the curves representing the last columns of Table IV.

It is necessary to note the fact that as n decreases, besides the influence of light of the central spot due to the brighter source, its first bright ring begins to influence the result curve of the distribution of illumination along a meridional line. This fact explains the decreasing of the critical distances for the n less than $n=0{\cdot}2$, and therefore the decreasing of the distances necessary for a resolution of two sources. At the same time, for these values of n ($\leq 0{\cdot}2$) the minimum of illumination on a meridional line is found, not between the geometrical images, but displaced away from the geometrical image of the fainter source (negative displacement); the second maximum is also displaced negatively.

When n has the value $n=0{\cdot}025$ and less—*i.e.*, the illumination at the centre (geometrical image) of the diffraction pattern due to the fainter source is approximately the same as the illumination at the first bright ring due to the brighter source or less than this, the first bright ring overlaps the light of the central spot of the second source and we have no critical distance for these values of n.

It is evident that for the resolution of two point-sources of unequal intensity we need greater distances than for the case $n=1{\cdot}0$. Thus if we consider, as is generally done, the distribution of illumination only along a meridional line, and assume 3 per cent. difference between the illumination at its second maximum and that at minimum as a practical limit for resolution, we should need the following distances for the different values of n:—

Table VI.

n.	D.	n.	D.
0·9	3·33	0·3	3·91
0·8	3·44	0·2	3·96
0·7	3·55	0·1	3·91
0·6	3·64	0·075	3·81
0·5	3·73	0·05	3·51
0·4	3·82		

In the case of two point-sources it is quite necessary to investigate the distribution of illumination not only along a meridional line but in a whole focal plane of an object-glass. For this purpose we must draw the lines

of equal illumination (isophotes) in a focal plane. In doing that, we can follow the method proposed by Nagaoka *. This method essentially consists in the following:—We take two points at a given distance, and consider them as the geometrical images of two sources; then draw the lines of equal illumination (circles, in a case of point sources) for each point-source. The position of the lines of equal illumination we find from the curve representing the general expression ($1\,a$) drawn in a large scale. Thus we obtain:—

Table VII.

Illumination.	Z.	Illumination.	Z.
1·00	0·00	0·45	1·73
0·95	0·45	·40	1·84
·90	0·64	·35	1·96
·85	0·80	·30	2·08
·80	0·93	·25	2·21
·75	1·06	·20	2·36
·70	1·18	·15	2·53
·65	1·29	·10	2·74
·60	1·40	·05	3·00
·55	1·50	0·00	3·83
0·50	1·61		

At the points of intersection of any two circles the sum illumination will be the sum of two. Joining the points of equal sum illumination, we get the curves of equal illumination.

For the case of two point-sources of equal intensity, we give here, as examples, four figures for the distances $D=3·83$ $D=3·2$, $D=3·0$, $D=2·9$ (Pl. II. figs. 4, 5, 6, 7). In all these figures O_1 and O_2 represent two geometrical images of the sources, M_1 and M_2 the positions of the maxima of illumination. Scale of the original drawing: each centimetre corresponds to $D=0·2$.

From these figures we see that the isophotes corresponding to central minimum and those of a higher illumination intersect a meridional line at four points (two points being between two maxima, *i. e.* each maximum is surrounded by

* Nagaoka, "Diffraction Phenomena in Focal Plane of a Telescope with circular aperture due to a Finite Source of Light," Journal of the College of Science, Imperial University, Tokio, vol. ix. p. 344 (1898), and Phil. Mag. xlv. p. 1 (1898).

these isophotes separately). The isophotes corresponding to a lower illumination surround two maxima together. The latter isophotes at their middle parts are convex (in respect to the central point in the diffraction pattern) with decreasing curvature as the distance between the sources increases. As the distance between the sources decreases, all isophotes become more and more of oval shape.

In the case of two *line* sources, we treat the conditions of their resolution only in respect to the least difference of illumination at its two maxima and that at point between them, just perceptible to the eye; but in this case the positions of the maxima and minima of illuminations are the straight lines: also by straight lines are represented all the other values of illumination.

A more complex case, as we see from Pl. II. figs. 4, 5, 6, 7, will be the case of two point-sources. The places of maxima of illumination in the diffraction pattern are two points. The places of all other values of illumination are the isophotes of a different form.

It seems more than probable that in some cases the eye cannot detect the difference between the illumination at its maximum and that at its minimum on a meridional line, but detects the depression in the middle part of the compound figure—*i.e.*, the depression in the middle parts of the isophote corresponding to the threshold sensibility of the eye.

It is possible that owing to this fact nearly all laboratory determinations of the resolving power of objectives were made with pairs or series of line objects (Foucault, Rayleigh, Bigourdan, Nutting).

At the distances less than the critical distance all the isophotes intersect a meridional line only in two points; in their middle parts they become concave in respect to the central point and the observer sees a compound figure as "oblong" till the eye refuses to detect the difference of diameters of this oblong figure, then the latter appears as quite round.

For two points of unequal intensity the isophotes may be found in the following manner:—We take two points at a given distance, considering them as the geometrical images of two sources; draw for the brighter point the lines of equal illumination as before. To find the positions of the lines of equal illumination (radii of the corresponding circles) for the fainter source, the illumination considered is divided by n and with the result as an argument the corresponding

distance of each line of equal illumination from the geometrical image of the fainter point is found from the curve representing the general expression (I. a). Then, joining as before the points of intersection with equal sum illumination, we obtain the isophote in question.

We give here (Pl. II. figs. 8–15) as examples some cases of distribution of illumination in a focal plane for the following values of n and D :—

n.	D.	%.	n.	D.	%.
0·7	3·6	0·95	0·4	4·0	0·90
—	3·55	0·97	—	3·83	0·97
0·5	3·73	0·97	0·3	3·91	0·97
—	3·6	1·00	—	3·83	1·00

It may be seen from these figures that the isophotes corresponding to minimum of illumination (on a meridional line) and those for greater illumination intersect a meridional line in four points, dividing themselves into two separate branches; at the distances less than the critical one, all the isophotes intersect a meridional line only in two points. In this case the isophotes corresponding to illumination at the second maximum and those of less illumination surround two maxima together; the isophotes corresponding to illumination greater than that at second maximum surround only the first maximum. As the distances exceed the critical one for the corresponding value of n, the depression of isophotes, which approximately corresponds to the position of a minimum of illumination on a meridional line, becomes of greater curvature.

As in the previous case of two points of equal brightness, it is more than probable that for some cases the eye cannot detect the difference between illumination at its maximum and that at its minimum, but detects the central depression: *i. e.*, it cannot detect the contrast of illumination but detects a form of compound figure.

To determine the conditions of resolution of two luminous points by the eye, it is necessary to know the threshold sensibility of the eye for such a case of diffraction pattern: *i. e.*, to know the value of isophote corresponding to the threshold sensibility and to know the contrast sensibility of the eye for the same pattern.

The above given results show that it is very difficult, if not impossible, to take two point-sources as test objects

for the accurate determination of resolving power of an objective.

In all considerations about the mutual attraction or repulsion of stars' images the displacements of maxima of illumination in a diffraction pattern must be taken into account.

The distance between two very close stars (very close components of a double star), especially of equal brightness, must be corrected for this displacement of maxima of illumination.

IV. *Physiological Limits to the Accuracy of Visual Observation and Measurement.* *By* H. Hartridge *.

Contents.

Introduction.

Definition of the Limits to Accuracy which are considered in this paper.

HITHERTO those factors which concern the accuracy of measurement that have been given much consideration are :—

1. The instrumental errors.
2. The mental bias of the observer.
3. The probable accuracy of the observations.

It would seem that there is a fourth factor which has not been fully recognized: namely, the effect of the physiological properties of the particular sense organ which is used as the connecting link between the instrument and the brain

* Communicated by Sir J. J. Thomson, O.M., F.R.S.

of the observer. In this paper the sense of sight will alone be considered, partly because visual determinations are the most commonly used; partly because the eye has been found as the result of experience to give values which are constant and accurate; and partly because the physiological basis of the errors to which visual estimations are liable is more firmly established than it is in the case of the other special sensations, *e. g.* touch, taste, or smell.

Section I.

Acuity of the Eye for Grating and Double Star Test Objects.

As is well known, there are two features which tend to limit the accuracy with which a given optical instrument can perform. The first is known as the resolving power of the instrument. This defines the limit imposed by the optical performance of its lens system. The second is the fineness of grain of the screen or surface on which that image is formed. This concerns the type of sensitive surface which is being used for recording the optical image, whether it be, *e. g.* a photographic plate or the retina of the eye. In the case of vision, both these factors enter into the consideration of the accuracy with which measurements of different types can be performed by the eye. We can thus treat the eye from the point of view of an optical instrument, and from the wave theory of light we can calculate what separation there must be between the two objects in order that their images should be clearly defined on the sensitive surface of the retina. For example, Abbé's formula * for the resolving power of the microscope is $r = \frac{f \times \lambda}{d}$. Assuming f the focal length of the eye to equal 15 mm., and the mean diameter of the pupil "d" to be 3 mm., and the mean wave-length of white light to be 5600 $\mu\mu$, then r is found to equal $2 \cdot 9\,\mu$. This means that the images of two objects formed on the retina must be separated by this amount in order that they shall be resolved by the eye. Alternatively, if Fraunhofer's formula † for the resolving power of the telescope be employed, and if a suitable correction be introduced for the fact that the pupil is circular, as pointed out in a previous paper ‡, a precisely similar formula is obtained and therefore

* Abbé, *cf.* Conrady, Journ. Roy. Micr. Soc. p. 610 (1904) and p. 541 (1905).

† Fraunhofer, *cf.* Airy, Camb. Phil. Trans. p. 283 (1834).

‡ Hartridge, Journ. Physiol. lii. p. 231 (1918).

a similar value for the resolving power. I find that Everett and Porter * have reached the same conclusion.

The histological limit to the resolving power of the eye is set by the distance separating two of the sensitive structures of the fovea. This is the part of the retina on which the images of objects are caused to fall when the greatest acuity of vision is required. The foveal cones as measured by a number of observers are summarized by Parsons † and are found to have diameters which vary between $2{\cdot}0\,\mu$ and $5{\cdot}4\,\mu$, the probable mean value being $3{\cdot}2\,\mu$. In order, therefore, for the images of two objects to be resolved separately by the eye, it is necessary that there should be at least one unstimulated cone between those two cones on which the images fall, that is, that the images must be separated by at least $3{\cdot}2\,\mu$. This histological limit will be seen to agree fairly closely with that obtained on purely optical grounds for white light and a 3 mm. pupil.

The calculation of the optical limit of the resolving power of the eye was made on the supposition that the image was a perfect one, that is, that the lens system of the eye was free from all types of aberration and conformed as a perfect lens system accurately to deductions made from the wave theory of light. This supposition is, of course, far from being the case, for the eye is known to suffer from considerable amounts of chromatic difference of focus and chromatic difference of magnification. Spherical aberration also enters to a greater or less extent, and in certain cases also astigmatism must be taken into account as well. For these reasons, therefore, we should expect the optical performance of the eye to fall far short of the theoretical performance that might be expected of it on the supposition that it behaved like a perfectly corrected lens system. It is surprising, therefore, to find that the limit to the resolving power obtained by actual experiments with the eye is not very different from that which a theoretically perfect lens system would give. In support of this statement the following values may be given :—In the case of a double star the limit is $4{\cdot}5\,\mu$. In the case of the grating test object it is $3{\cdot}6$–$4{\cdot}6\,\mu$. In the case of the chequer test object it is $4{\cdot}6\,\mu$. It will be seen that limits to the resolving power obtained by using test objects of various types, give limits which vary between $3{\cdot}6\,\mu$ and $4{\cdot}6\,\mu$, these being the distances separating on the retina the geometrical images of the separate elements of the test objects in question. The mean of these values,

* Everett and Porter, Journ. Roy. Micro. Soc. p. 405 (1907).

† Parsons, 'Colour Vision,' 1st edition, p. 9 (1915).

namely 4·2 μ, is not very different from that which would be expected either from the histological limit (3·2 μ) set by the diameter of the foveal cone, or the optical limit (2·9 μ) obtained for the eye regarding it as a perfect lens system. The problem as to how it is that the eye can behave in such a satisfactory way, in spite of the aberrations from which it is known to suffer, has already been considered in some detail in a previous paper *. With regard to the effects of these aberrations on visual acuity, it is interesting to note that while Helmholtz † found no marked improvement in the visual acuity on compensating the chromatic aberration of the eye by means of a chromatically overcorrected focal flint and crown glass lens combination, Luckiesh ‡, on the other hand, by using monochromatic light of considerable intensity obtained from a mercury vapour lamp, has obtained an improvement of roughly 15 per cent., on substituting pure green light for daylight. Of the two experiments I would regard Luckiesh's as the more reliable, because any additional optical system, particularly one containing a combination of highly curved lenses, must introduce other aberrations at the same time that it corrects the chromatic aberration in the eye. If this 15 per cent. improvement for monochromatic light be applied to the mean acuity value obtained for white light for the grating test object (4·2 μ), a value of 3·6 μ is obtained for the distance separating the centres of the images of the bright bars of the grating on the retina. This value for visual acuity by monochromatic light is, as might be expected, even closer to the limit set by the diameters of the cones (3·2 μ) and by diffraction (2·9 μ) than is the experimental value obtained for white light (4·2 μ). Two other factors of importance now require to be mentioned, namely, the effect of light intensity and pupil diameter. With regard to the former, it is generally agreed § as the intensity of illumination of the test object increases, that an increase occurs in the acuity of the eye. Very marked is the increase at first, but with increase of light intensity a level is reached at which the acuity stays constant in spite of increased illumination. It is probable that beyond this point an increase in illumination is accompanied by decreased visual acuity, but further experiments are required to set this on a firm basis.

With regard to pupil diameter, it has been shown by

* Hartridge, Journ. Physiol. lii. p. 231 (1918).
† Helmholtz, Physiol. Optics, 3rd edition, p. 157.
‡ Luckiesh, Electrical World, lviii. p. 450 (1911).
§ Parsons, Roy. Lond. Ophth. Hosp. Report xix. Part i. p. 109.

Lister * and Cobb † that between certain limits (3 mm. and 5 mm. pupil diameter) visual acuity is not appreciably affected by the size of the pupil. Below 3 mm. both observers agree that acuity decreases with a decrease in pupil diameter. Above 5 mm. Cobb finds a decrease, while Lister finds practically the same value as at 3·3 mm. We may say then, so long as the eye be in a state of adaptation, no bright source of light be shining into it, and the pupil diameter be approximately 3 mm., that the maximum theoretical resolving power is very nearly attained.

It should, however, be carefully noted with regard to these experimental limits that they have been obtained as the result of tests carried out by skilled observers using prolonged observation under specially favourable conditions of lighting, etc., and that such good values would hardly ever be obtained even by keen-sighted observers for ordinary conditions of lighting in everyday life. We may say, then, that whereas the eye actually *can* under specially favourable conditions separate the images of objects when they are separated at the retina by $3\,\mu$ to $4\,\mu$, it would greatly facilitate observations if they were not separated by less than, say, five times this value. It is necessary to point this out, because rules have been made by Helmholtz and others for the limiting useful magnifying power of the microscope when objectives of different numerical aperture are used ‡. The use of very low magnifications has been advised on the ground that the eye is well capable of resolving such a fineness of structure. It should be pointed out, however, that although it is possible for the eye under almost ideal conditions to resolve a certain degree of fineness, yet it cannot do so unless these ideal conditions are attained, and, further, even under these circumstances the fatigue caused by the process of observation would be found to be very great. It should be noted, moreover, that the grating test object used for the determination of visual acuity consists of metal wires placed in front of a brightly-lit white surface. The bars are absolutely black, and the clear parts between quite bright. In the case of a periodic structure seen under the microscope this is not the case, because no matter how perfectly the instrument is corrected there is a certain

* Lister, *cf.* Conrady, Journ. Roy. Micr. Soc. xxxii. p. 41 (1913).

† Cobb, Am. Journ. Physiol. xxix. p. 76 (1911), and Psychol. Review, xx. p. 425 (1913).

‡ Helmholtz, *Wissenschaftliche Abhandlungen*, vol. ii. p. 185, also Pogg. *Ann.* Jubelband, p. 569 (1874); Mallock, Nature, cix. p. 205 (1922).

amount of haze which makes the black parts milky. Aberrations, diffraction, reflected and scattered light all play their part in producing this haze. Further, it becomes relatively more and more marked as the structure becomes sufficiently fine to be reaching the limit to the resolving power of the instrument. When such images are formed on the retina the conditions are very different from those of the grating test object described above.

Experiment shows that as the difference between the amount of light reaching the eye from unit angle of the black and white parts of the image is lessened, the lower is the value of the visual acuity reached. The following values obtained by Aubert may be quoted as examples :—

Distances in μ between the centres of the retinal images of two white squares in order that they should be resolved.

Size of square in μ	8·3	6·6	5·5	4·7	4·1	3·7
Distances for black ground	10·3	11·0	12·2	14·9	15·8	18·5
" " grey "	10·8	11·5	12·6	14·5	19·4	23·2

A similar decrease in visual acuity is obtained if under the same intensity of illumination grey letters on a white ground are substituted for black letters on a white ground. We conclude, then, that the image of a grating formed on the retina will require to be more highly magnified if the black parts of it are milky than is the case if a perfectly black and white image is formed. This is, then, an additional reason for requiring that the magnifying power used in such an instrument as the microscope should be several times greater than the theoretical minimum. Experiment will show roughly how much additional magnifying power is required. Thus the diatom *Pleurosigma angulatum*, well known to microscopists as an object for testing dry lenses of medium power, has a regular periodic structure the approximate distance between the elements of which is ·58 μ. This diatom placed 250 mm. from the eye would give images on the retina, the distances between which would be ·035 μ. But experiment shows that under ideal conditions the eye can resolve images 4·2 μ apart. A magnification of 120 diameters would be sufficient to cause the diatom markings to be this distance apart on the retina. This magnification should therefore, theoretically, be sufficient to make them visible. Now, since the formula for the resolving power of the microscope is $R = \frac{\lambda}{2\,\mathrm{N.A.}}$, then if white light is in use,

$\lambda = 5800$ A.U. approximately, and the N.A. required will be found to equal 0·5 almost exactly. Helmholtz's rule for the magnification, namely, $264 \cdot 5 \times$ N.A., should therefore give a value of 133. That is roughly equal to the lowest possible value found above, namely 120. Now experiment shows that to the author's eye a magnification of 500 is near the mark. That is, Helmholtz's rule is approximately 4 times too small even for well-marked structures. The distance separating the images on the retina at the magnification of 500 required by the author's eye becomes $17 \cdot 5\,\mu$, which corresponds roughly to 5 times the diameter of a cone ($3 \cdot 2\,\mu$). Therefore, if the geometrical images of two neighbouring parts of the periodic structure fall on two cones, there will be between these cones on the average three other cones which will correspond to the interval between the periodic structures.

It should be pointed out that this apparently high degree of magnification required by the author's eye is not due to abnormality in its refraction. Experiment shows * his eye to be emmetropic with a visual acuity on Snellen's scale of 6/3, *i.e.* approximately twice that of Snellen's normal value. Persons with normal vision attaining 6/6 on Snellen's scale would therefore presumably require a higher magnification still, possibly 1000 (at 0·5 N.A.) for clear vision, and persons with abnormal sight a corresponding increase in the magnifying power. It is obvious from this consideration how erroneous is Helmholtz's estimate of 133 diameters for 0·5 N.A. (*i. e.* $266 \times$ N.A.). Gordon † reaches a similar conclusion.

Abbé's rule, viz. $1000 \times$ N.A., is nearer the mark, but this would certainly be too low for persons with normal sight under adverse conditions and for persons with abnormal sight under almost all conditions. A safer rule would be $\dfrac{2000 \times \text{N.A.}}{\text{Snellen's standard of visual acuity}}$. A person with very good vision reaching Snellen's standard 6/3 would thus require $\dfrac{2000 \times \text{N.A.} \times 3}{6} = 1000 \times$ N.A. (which is equal to Abbé's rule). A person with 6/9 vision would require $3000 \times$ N.A., and so on.

The conclusion to be drawn from this section is that optical, histological, and experimental considerations give to

* Hartridge and Owen, Brit. Journ. Ophthalmology, Dec. 1922.

† Gordon, Journ. Roy. Micro. Soc. 1903, p. 418.

the eye a limiting resolving power of about 4 μ at the retina, which value corresponds to slightly less than an angle at the eye of 1 minute of arc. It is not surprising, then, that this limit has been accepted for the purposes of calculating the numerical values of such properties of the eye as depth of focus. Helmholtz, for example, used this value for calculating what he considered to be the greatest attainable acuity of stereoscopic vision, and in consequence it was not generally recognized for many years after that experimental determinations of stereoscopic vision showed that the eye had a very much greater stereoscopic acuity than Helmholtz's values gave.

Section II.

Acuity of the Eye for Contours under Various Conditions.

Volkmann appears to have been the first, in 1863, to point out that the acuity of the eye for contours can greatly exceed this value of 4 μ displacement at the retina. He determined the smallest differences between the breadth of two white contours on a black ground that can be perceived by the eye. He found this difference stated in angular measure to be as little as 10 seconds of arc (0·73 μ at retina). Wulfling, in 1892, confirmed Volkmann's observations and obtained values of 10–12 seconds of arc (mean = 0·79 μ at retina). Hering, in 1899, showed that the edges of two rectangles did not appear in line with one another if there was a relative displacement of approximately 10 seconds of arc (0·73 μ at retina). I have recently confirmed Hering's value by means of a test object, consisting of a white circle some 5 cm. in diameter, having on one part of its circumference, 10 mm. in length, a small projection 2 mm. in width. This disk was mounted in front of a sheet of black paper, so that the projection could be rotated into different meridians about a pin passed through the centre. The projection was placed in different meridians by an assistant and was observed at increasing distances until the point was reached at which the projection was correctly observed eight times out of ten. The distance was found to be 130 feet, and at this distance the projection was calculated to subtend 11 seconds of arc at the eye of the observer (0·79 μ at retina). This confirms Hering's value almost exactly.

It is clear from these experiments that under certain

circumstances the acuity of the eye greatly exceeds 4 μ displacement of the retinal image. Other observations on the accuracy with which contours can be set into coincidence confirm this view. For black lines on a white ground Bryan and Baker * obtained an accuracy of 12 seconds of arc (0·87 μ on the retina). Hartridge obtained 8·5 seconds (0·62 μ), Stratton † obtained 7 seconds (0·51 μ). For white lines on a black ground Bryan and Baker obtained 9·5 seconds (0·69 μ) and for a split line 8·0 seconds (0·58 μ). It will be seen that the mean accuracy of the coincidence method is 9·3 seconds of arc, *i. e.* a displacement of the retinal image of 0·67 μ.

It is found, further, that the acuity of stereoscopic vision provides yet another case of the accuracy of perception of contours. Thus the values below demonstrate a mean accuracy of 8·2 seconds of arc (0·57 μ on the retina). Pulfreich obtained a value of 10 seconds (0·7 μ on the retina). Heine obtained values varying between 6 and 13 seconds (0·4–0·9 μ), Bourdon obtained 5 seconds (0·35 μ), Crawley 3 seconds (0·2 μ), Breton ‡, in the case of one skilled observer, found an accuracy of 4 seconds (0·28 μ).

The visibility of single black contours on a white ground, or white contours on a black ground, provides another example. Aubert found a black line on a white ground visible to the eye when its edges subtended an angle of 6 seconds of arc. Smith Kastner obtained a value of 3·5. Hartridge, using bright brass wire, found 3·6, and Hartridge and Owen detected the direction taken by the arms of an L (bent out of black brass wire and suspended in front of a white background so that no shadow was cast) 8 times correctly in 10 trials when its edges subtended 3·1 seconds of arc at the eye. (A decrease in the angle of 5 per cent. resulted in 6 mistakes being made in 10 trials, showing how abruptly the limit is reached.) The mean value obtained is 4 seconds of arc (=0·29 μ at retina). In the case of white contours even lower values have been obtained, in fact any white object, no matter how narrow, is visible if the intensity of illumination is sufficiently great. Seidentopf's experiments on the ultra microscope have amply demonstrated the truth of this statement.

* Bryan and Baker, Proc. Opt. Convention, 1912, p. 252.
† Stratton, Psychol. Review, vii. p. 429 (1900).
‡ Breton, Journ. Roy. Navy Med. Section, vi. p. 288 (1920).

These different mean values may for convenience be summarized in Table I. below.

TABLE I.

Visual Acuity of the Eye for Contours.

	Seconds of Arc.	At Retina.
Visibility of a black contour on a bright ground	4·0	0·29 μ
Appreciation of the position of a contour	10·5	0·76 μ
Accuracy of coincidence of two contours	9·3	0·67 μ
Acuity of stereoscopic vision	8·2	0·57 μ
Mean	8·0	0·57 μ

It will be seen that, taking 8·0 seconds as the mean accuracy of the above determination, the acuity of the eye where contours are concerned is roughly 7 times as great as that to be expected on the histological and optical grounds mentioned in Section I. We have therefore to consider how this accuracy is to be reconciled with the calculated resolving power of the eye, and the finite diameter of the cone. With regard to the former, Rayleigh * gave the following formula for calculating the least width of black line for visibility when viewed against a self-luminous background :—

$$\frac{I_1 - I_0}{I_0} = \frac{2\alpha}{\pi}.$$

Perceptible percentage decrease of illumination multiplied by angle subtended at the eye by the lines of a grating for resolution to occur, equals angle subtended at eye by just visible black line. He showed, by experiment, that

$$\frac{2\alpha}{\pi} = \frac{1}{9{\cdot}1}, \quad \therefore \quad \frac{I_1 - I_0}{I_0} = 11 \text{ per cent.}$$

Taking, similarly, objects illuminated by borrowed light, he showed that the formula is

$$\frac{I_1 - I_0}{I_0} = \frac{4\alpha}{\pi}.$$

Perceptible percentage decrease of illumination multiplied by angle subtended at the eye by the lines of a grating for

* Rayleigh, Journ. Roy. Micr. Soc. 1903, p. 474.

resolution to occur, equals angle subtended at eye by just visible black line.

If these α values for self-luminous gratings and those illuminated by borrowed light be compared, it will be seen that a self-luminous grating is twice as easily resolved by the eye as one illuminated by borrowed light. But that the converse is the case for resolution of a single narrow black line. Since then the angle of the black line is roughly one-ninth that of the bars of the grating when both are on self-luminous grounds, the ratio must be one to thirty-six, when borrowed light is used. Now the experimental values for the grating test object thus illuminated have already been given above, the mean being 58 seconds of arc. A black line should therefore be visible when it subtends an angle one thirty-sixth of this, viz., 1·60 seconds of arc. But the lowest value given above was 3·1 seconds of arc, which is roughly twice Rayleigh's value. If the eye could be fully corrected for aberrations it is probable that a nearer value would be obtained to 1·60 seconds of arc. It is therefore clear that the performance of the eye does not exceed the limits calculated according to the diffraction theory of light.

We have now to consider how the limit apparently set by the diameter of the foveal cones comes to be exceeded :—

Since it has never been found possible to analyze experimentally the structure of the image formed on the retina, calculations based on the amounts of aberration present in the eye have to be depended on to give the intensity from point to point of that image. It is found in this way * that the image, even of a point source, spreads itself over a large number of cones, but the intensity falling on the various cones will be very different. The two factors of principal importance are those mentioned in Section I., namely, diffraction and chromatic aberration †. I have also mentioned that increase in pupil diameter increases one but decreases the other, thus explaining how it is that the performance of the eye stays constant for pupils between 3 and 5 mm. Calculation of the effects of these two factors on the image of a point source shows that the cone receiving the strongest stimulus is the one on which the centre of the pattern falls. The surrounding cones will be less strongly stimulated. A just noticeable white object will be that one which just stimulates the cone on which the centre of the image falls by an amount appreciably greater than that by

* Hartridge, Journ. Physiol. lvii. p. 52 (1922).

† Hartridge, Journ. Physiol. lii. p. 176 (1918).

which it stimulates surrounding cones. A just noticeable black object will be that causing an appreciably less strong stimulus, and so on. For the movement of a contour to be perceived it must cause a cone on one side of the edge of the image to receive an appreciably stronger stimulus, and that on the other an appreciably weaker one, than before. The acuity of the eye will therefore depend less on the diameter of the cones than on its ability to perceive small changes in light intensity. If from various experimental values the amount of this just perceptible intensity difference be calculated the following values given in Table II. are obtained.

Table II.

Difference Threshold for Light Intensity required in different circumstances to account for the Acuity of the Eye.

Single black line on white ground	13 per cent.
Grating	20 ,, ,,
Double line	12 ,, ,,
Stereoscopic vision	12 ,, ,,
Coincidence of two contours	22 ,, ,,

It will be seen that on the average there must be a difference of 16 per cent. between the light intensity falling on one cone and that falling on its neighbour for that difference to be perceived. It should be noted that 16 per cent. is a threshold value very much higher than that found by experiment to be required for large groups of cones. Here 0·5 or 0·6 per cent. suffices. Thus it was found by Helmholtz that if there is a difference of more than one part in 165 in the illumination of two contiguous areas, the difference in illumination will be perceived. Further, it was found by the author that two absorption bands 100 A.U. wide can be set into coincidence with an error not greater than 1 A.U.* Spectro-photometric measurements † show that 2 A.U. change of wave-length produces a 1 per cent. change in intensity in that part of the band where the density change is greatest. For a wave-length change of 1 A.U. to be perceived the retinal cones must be capable of observing a 0·5 per cent. change in intensity. Various factors affect the acuity of the eye for perceiving differences of intensity, one is the intensity of the light and the other the adaptation of the eye.

* Hartridge, Proc. Roy. Soc. (Now in course of publication.)
† Hartridge, Journ. Physiol. xliv. p. 1 (1912).

It may be sufficient to summarize briefly the data concerning these factors as follows. If the light is not bright enough acuity is low, and probably if the light is too bright also acuity is low, *i. e.* there is between the two extremes an "optimum" brightness *. With regard to adaptation, both incomplete and excessive adaptation appear to be disadvantages, *i. e.* there is an optimum for adaptation also.

Section III.

Linear Measurements by the Method of Coincidences.

The various types of linear measurement which are employed in practice may be divided under two headings: those which use the principle of coincidence and those which use the principle of interpolation. As an example of the first, may be given the measurement of an object by a scale and vernier: as an example of the second, may be given the measurement of a length by means of an ordinary millimetre scale. The fractions of a millimetre are determined in the first method by means of a mechanical contrivance, and in the second method by means of imaginary subdivisions of the millimetre into tenths (and possibly into hundredths) by the eye. For convenience, these two types of measurement will be considered separately, *measurements by means of coincidences* being considered in this section.

Baker and Bryan † appear to have been the first to make direct measurements of the accuracy with which one line can be set into coincidence with another. They investigated three cases: (1) that of a black line on a white surface, (2) that of a white line on a black surface, and (3) that of a black line the centre part of which was adjustable, which may be referred to as a "split" line. Some of the values which they obtained have been given above in Section II. It will be seen that the average error of coincidence methods may be as low as $0{\cdot}5\,\mu$ on the retina, *i. e.* $8{\cdot}3\,\mu$ at 25 cm. from the eye.

It may at this point be interesting to consider what effect on accuracy the want of distinctness in the object will have: for example, absorption bands produced by pigments in the spectrum are found to have blurred outlines, which makes the measurement of the apparent boundaries of their edges a matter of very great difficulty. It is also difficult, owing to

* Parsons, Roy. Lond. Ophth. Hosp. Reports, xix. Part i. p. 110.
† Bryan and Baker, Proc. Opt. Convention, 1912, p. 252.

the apparent considerable width of the band, to make accurate observations of the apparent centre of the absorption. When, however, the method of coincidences is employed, two such similar bands being set in line with one another, then it is found that the accuracy of measurement is very greatly increased. Thus, without the method of coincidences it is difficult to obtain an accuracy of less than 10 A.U., whereas with the method of coincidences it is found that an accuracy of 1 A.U. is not difficult to obtain *. It is possible to convert this accuracy of 1 A.U. into values similar to those previously given, *i. e.* for the apparent accuracy of setting the images on the retina into coincidence. The value thus obtained is found to be $0{\cdot}76\,\mu$, which is but little greater than the accuracy of coincidences found by Bryan and Baker in the case of a sharply-defined black and white line. The reason for this apparently small difference between the values for sharp and blurred lines is that, in spite of the fact that a mathematically sharp line is measured on the one hand, and an absorption band with indefinite edges on the other, there is not so marked a difference in the two sets of images formed on the retina, because chromatic aberration and diffraction operate together to give to the image, both of a mathematically sharp line and an absorption band, an indefinite outline. Two important physiological factors operate together in both cases to reduce this apparent blurring in the retinal image, these are (*a*) the value of the absolute threshold of stimulation of the fovea for vision and (*b*) simultaneous contrast. With regard to the former, it is found by experiment that with a particular degree of adaptation of the retina there is a certain luminosity below which all objects appear black, and, alternatively, there is a certain threshold above which all objects appear white, but the effects of this upper threshold are not so important in the cases under consideration as the lower one. With regard to simultaneous contrast, it is found by experiment that the differences between the stimuli applied to any two neighbouring areas of the retina are greatly increased †, so that, for example, a grey surface on a black ground appears to become white, and so on. This factor must have very important effects on the appearance of what may be called gradients of intensity, that is, in cases like absorption bands in which there is a more or less gradual change of the intensity of illumination as one passes from one part of the image to another.

* Hartridge, Proc. Roy. Soc. (Now in course of publication.)
† Hartridge, Journ. Physiol. l. p. 47 (1915).

Simultaneous contrast has the effect of making all such gradients appear very much steeper than they actually are, and, if there should be any discontinuities in the intensity curve, of accentuating such discontinuities so as to cause the areas bounding such discontinuities to appear still more different in intensity. It is these physiological factors which cause images on the retina to appear very much sharper than the theoretical distribution of intensity would lead one to suppose would be the case. These factors both reduce the blurring introduced into the retinal images by the aberrations in the lens system of the eye.

These same factors operate in the case of objects which themselves have blurred outlines, *e. g.* absorption bands, thus causing the accuracy with which two such bands can be placed in coincidence to be not very different from that obtainable in the case of objects the edges of which are sharp.

We can draw from the above considerations the following practical conclusions :—

The accuracy of the method of coincidences in the case of both sharp and blurred lines (absorption bands, etc.), is very great, being between 4 and 5 times the resolving power of the eye. The retinal images are adjusted into coincidence with an average error of less than 0·8 μ. This corresponds to an angle of 11 seconds, or to one-eightieth part of a millimetre at 250 mm. distance from the eye (*i. e.* the ordinary distance at which a scale or instrument would be placed for clear vision with the unaided eye). It should therefore be possible with an accurately constructed vernier to subdivide the millimetre into 80 parts. For this purpose 40 vernier lines per millimetre would be required at least, since apparent coincidence with two neighbouring lines would be construed as coincidence with a line not actually present but midway between the two. In practice, a $\frac{1}{2}$ mm. scale is more suitable, and this will require but 20 vernier lines to obtain the full accuracy available. If the scale is viewed through a magnifying-glass a corresponding change in scale and vernier should be made. Further, any considerable break in a contour of regular form will be identified if it subtends at the eye an angle of more than 11 seconds. For example, through a telescope the satellites of one of the planets should be recognizable if the magnified image has an excrescence due to the satellite of more than 11 seconds. In the case of Venus, observers at Greenwich claimed to have seen the satellite which subtends an angle of 0·5 seconds of arc. This through a telescope of 1000 diameters should equal 50″,

i. e. nearly 5 times the limiting value. There is little doubt that this satellite should be visible under good atmospheric conditions.

The coincidence method of measurement is liable to these types of error :—

(*a*) Line thickness parallax.
(*b*) Malfocus parallax.
(*c*) Colour parallax.
(*d*) Intensity and visual threshold parallax.

(a) *Line thickness parallax.*—In the case of indices which are attached to scales which lie side by side (*e. g.* scale and vernier), the accuracy is not so noticeably affected by line thickness as by a difference in thickness of the lines on the two scales. But even in the latter case the inaccuracy is small compared with that introduced into interpolation measurements by thick lines. (This will be considered later.)

(b) *Malfocus parallax.*—This is the error familiar to astronomers when two points under observation differ in position in the direction of the line of sight, or when two scales which are being placed in coincidence with one another are in different planes relative to the eye, *i. e.* one scale being nearer to the observer than the other scale. The avoidance of error due to these causes is, as a rule, simple, and practical details are well known. It should be noticed, however, that when both eyes are being used together during observations, it is usually one eye—called by oculists the master eye—by means of which the settings are made. It is known, however, that in certain individuals neither eye is master, but one eye is sometimes used for setting and, on other occasions, the other eye. This selection takes place without the observer being aware that any change is taking place. Errors may be introduced from this cause if the two scales do not lie in the same plane.

(c) *Colour parallax.*—Donders pointed out that the eye suffers from chromatic difference of magnification ; Gullstrand estimates that violet images are 3 per cent. smaller than red ones. I have confirmed this value *, and find that since the fovea is approximately 0·6 mm. to the temporal side of the optical centre of the retina, the foci for different coloured rays fall approximately along a straight line which

* Hartridge, Journ. Physiol. lii. p. 176 (1918).

cuts the optic axis 11·5 mm. in front of the retina. The centres of the aberration disks formed on the retina by rays of different colour do not therefore coincide, but the blue disks fall relative to those of the yellow on the nasal side and red disks on the temporal. To this is due the phenomenon of chromatic stereoscopy described by Donders. The relative displacements of the centres of red and blue images on the retina is approximately 0·02 mm. That is an apparent displacement of about 0·3 mm. in a scale illuminated by monochromatic red light in reference to a scale lit by pure blue-violet light placed side by side with it at, say, 250 mm. distance from the eye of the observer. Two examples will be given :

Case 1. Suppose that first one eye and then the other be used for measurement, an error of 0·6 mm. might occur if the scales were illuminated by pure red and blue light respectively. Scales tinted red and blue would suffer smaller apparent relative displacements, because these would not reflect pure spectral colours differing greatly in refrangibility, but only mixed rays the mean refrangibility of which would be less. A smaller but still noticeable error should occur if, say, one scale be of gold and the other of silver (a common arrangement for a scale and its vernier).

Case 2. Suppose that the length of an object part blue and part red is being measured at 250 mm. from the eye ; then an error in the length of 0·3 mm. might result.

A number of other examples might be given, but errors of this nature can be avoided by avoiding coloured scales, making corresponding scales of similar metals, and measuring all coloured objects under monochromatic light of specified wave-length.

(d) *Intensity and visual threshold parallax.*—It has been found by experiment that the measurements of the mean wave-length of absorption bands are influenced

(*a*) by the relative brightness of the illumination on the two sides of the bands (the band appears to be shifted towards the darker side), and

(*b*) by the relative sensitiveness of the two parts of the retina on which the images of the edges fall * (the band appears to be shifted to that side falling on the less sensitive half of the retina).

Since the image on the retina, even of a mathematical line,

* Hartridge, Proc. Roy. Soc. B. lxxxvi. p. 128 (1913).

is a diffuse one, such a line must show similar though probably smaller shifts under the influence of these two factors. Suppose, for example, that a scale, consisting of black lines on a white ground, be discoloured by a localised dark stain, then if the edge of the stain happens to coincide with one of the lines, so that on one side of the line the ground is darker than it is on the other, then it would be expected that measurements made in reference to such a line would be inclined towards the darker side. These effects of intensity are probably negligible under ordinary circumstances. It would, however, appear best in practice to avoid any possible errors due to them by the use of clean scales.

The retinal factor is not so readily disposed of. It would seem best in practice either to examine the scales with their ends alternatively to the right and left, or to perform coincidence settings alternatively with the scales on which the determinations are being made and with another set as nearly like them as possible which can be examined with their ends alternatively to right and left.

The four cases just considered apply to the unaided eye, and, with suitable modification in the actual values of the errors, to the eye when looking through an optical instrument which has an exit pupil large in comparison with the pupil of the eye (*e. g.* a Galilean telescopic system or a magnifying-glass). When the exit pupil of the instrument is small in comparison with the pupil of the eye, it is found that three sources of error may occur in instruments which employ the coincidence method of measurement:—

i. The first is well known, namely, when two objects to be set into coincidence do not lie approximately in the same plane, for if the eye suffers from spherical aberration (as is almost always the case), then a small side shift may be introduced, causing the one image to move in reference to the other when the eye is moved from side to side.

ii. In instruments like the direct-vision spectroscope, a side shift may be introduced if the colours of the two images to be brought into coincidence are not the same; for example, in the direct-vision spectroscope it is usual to find a scale illuminated by white light in juxtaposition to the spectrum to be measured. Suppose, for example, that there are two lines in the spectrum under observation, one in the red and one in the violet. It will be found, owing to chromatic aberration, that if the eye was accurately

focussed for white light, rays from the violet part of the spectrum would be brought to a focus in front of the retina, whereas rays from the red end of the spectrum would be brought to a focus behind the retina. If the exit pupil of the instrument accurately coincided with the optical axis of the eye, no deflexion of these images relative to one another would result: but if the exit pupil of the instrument was to the left side of the optical axis of the eye, then it would be found that violet images would be displaced to the left and red images to the right, and *vice versa*. Supposing, then, the spectrum has its violet end on the left and its red end on the right, the difference in wave-length of the red and violet lines would appear to be exaggerated so that the violet line would appear to have too small a wave-length and the red line one that was too large. Alternatively, if the exit pupil of the instrument were to the right of the optic axis of the eye, the reverse conditions would hold good, and the converse would be the case. It was found in the case of a direct reading spectroscope by Zeiss that an error of approximately 50 A.U. in the wave-length measurements of the violet helium line ($\lambda = 4713$) could be introduced by causing the rays from the instrument to pass through the extreme right or left-hand edge of the pupil.

iii. In cases where an instrument with split field is being used, it is found that a relative shift of the two fields can occur if the exit pupils of the rays which have illuminated the two fields do not accurately coincide with one another, and if the eye suffers from spherical aberration (as is almost always the case). For example, supposing the upper of two fields to have its exit pupil to the right of the exit pupil in the lower of the two fields, and supposing that the exit pupil of the upper field passes through the periphery of the lens system of the eye (which is usually found to suffer from spherical under correction), then the images in the upper field will be found to be displaced to the right in relationship with images occurring in the lower field, and therefore errors will be introduced. This method is used in a large number of instruments, *e.g.* the chemical burette and the slide-rule.

Section IV.

Interpolation Methods of Measurement.

This method depends on the visual subdivision of a space bounded by two lines. Suppose, for example, that the position of a line is being measured by means of a millimetre scale: the millimetre is subdivided by the eye into tenths and possibly hundredths, and the number of such tenths or hundredths added on to the whole number of millimetre divisions which the object is found to measure. The accuracy with which such a subdivision can be effected has been made the subject of measurement. In Table III. are given the mean values of ten measurements of certain distances; in the first column as adjusted by interpolation by the eye, and in the second column as actually measured by means of the vernier. In the third column are given the differences between these two methods of measurement.

Table III.

Showing Errors in Judging the Subdivisions of a 1 mm. Scale into Tenths at 25 cm. from the Eye.

By Eye.	By Vernier.	Error in μ on Scale.	Error in μ on the Retina.
0	0·000 $\pm$0·015	$\pm$0·015	$\pm$0·9
+0·1	+0·196	+0·096	+5·8
+0·2	+0·286	+0·086	+5·2
+0·3	+0·373	+0·073	+4·4
+0·4	+0·462	+0·062	+3·7
+0·5	+0·535	+0·035	+2·1
+0·6	+0·635	+0·035	+2·1
+0·7	+0·666	−0·034	−2·1
+0·8	+0·727	−0·073	−4·4
+0·9	+0·829	−0·071	−4·3
+1·0	+1·000 $\pm$0·015	$\pm$0·015	$\pm$0·9

An observation of this column will show that the whole reading in millimetres is very accurately judged, this depending on the coincidence of two lines; but the readings on other parts of the scale will be found to show considerable inaccuracy: *e. g.* the bisection of the space is judged inaccurately by 2 μ, the subdivision into the first tenth is judged inaccurately to 6 μ, and 4 μ for the two ends of the division respectively, other values giving errors which vary between

these two values. It will be seen, therefore, that there is considerable inaccuracy in making this type of measurement, an inaccuracy which was not less than $30\,\mu$ at 25 cm. from the eye, and was in one case more than $90\,\mu$, whereas the corresponding errors found for coincidence measurement at 25 cm. from the eye are very much smaller, being about $10\,\mu$. This inaccuracy, which appears to be large everywhere and to increase as one of the lines is approached, seems to be due to a physiological factor, namely, that in subdividing a space bounded between two lines, it is not the space as measured from the centre of one line to the centre of the other line which is subdivided by judgment, but, for example, in the case of black lines the white area which extends from the edge of one black line to the edge of the other. This means that the area subjected to subdivision is not the whole area, but is the area from which the thickness of the lines has been subtracted. It would therefore be expected that this physiological error would be the greater the thicker the lines bounding the space, and I have a certain amount of evidence that that is the case. In the subdivision, therefore, of a scale by means of the method of interpolation, it is essential that the line thickness be kept as small as possible, and, further, if the personal error similar to that shown in Table IV. above be previously ascertained by the observer, a correction should be applied for the measurements made by the method of interpolation. The effects of colour if it be different in the two halves of the field, and of malfocus and want of coincidence in the exit pupils of the optic systems of the two fields, are similar to those considered above in the case of linear measurements.

Section V.

Contact Methods of Measurement.

In this method the images of two objects formed on the retina are gradually caused to approach until they just appear to come into contact with one another. For example, this method is employed when the size of the image of a microscopic object is being measured by means of the spider-line micrometer eyepiece. Another example is the case of the sextant, in which the bright image of the sun is brought into contact with the apparent edge of the horizon or with another image of the sun obtained by reflexion. Bryan and Baker *, who made a very careful

* Bryan and Baker, Proc. Opt. Convention, 1912, p. 252.

survey of the accuracy of contact measurements obtained with different forms of object, found that the accuracy depended largely on the test object. In certain cases, adjustment stopped short of contact, whereas in other cases there was an actual overlapping of the images before the contact appeared to be obtained. In general, it would seem from their determinations that black objects on a bright ground are not quite brought into contact, *i.e.* the measurements are too large; whereas bright objects on a black ground are caused to slightly overlap, *i.e.* the measurements are too small.

These constitute errors in the contact method of measurement, which are unquestionably due to the physiological properties of the retina.

Now, Nelson * has published a formula by which the measurements of microscopic objects obtained by the contact method should be corrected so as to give the true dimensions of those objects. The formula he gives is :

$$\text{Correction in } \mu = \frac{\lambda}{5{\cdot}47 \times \text{W.A.}},$$

where λ is the wave-length of the light and W.A. = working aperture of the objective.

The values given by it must be added to the experimental value in the case of black objects and subtracted from the experimental value in the case of bright ones.

If we applied such a correction to Bryan and Baker's results, we should find that the experimental value for the dimensions of a black object which are already too large owing to the fact that "contact" appears to be obtained when the spider-line and the black object are still separated by an interspace, would be made larger still, *i.e.* that Nelson's correction would add to the existing error, whereas in the case of bright objects on a black ground the converse would be the case.

Nelson's formula can be tested in another manner, with a 3 mm. pupil the W.A. of the eye is 0·1 N.A., and when this is inserted into Nelson's formula (λ being 5600 A.U. for white light) the value 1·025 is obtained as the correction. This is equal to a visual angle of 14 seconds of arc approximately. White objects should appear too large by this amount, while dark objects should appear too small. Aubert has given experimental values obtained by varying the

* Nelson, Journ. Roy. Micr. Soc. 1903, p. 579.

distance between white lines on a black ground until the width of the interspace between the lines appeared equal to the width of the lines themselves. In this way he found that white lines subtending at the eye angles varying between 45 and 10 seconds of arc, appeared between 50 and 77 seconds of arc too large. Nelson's correction of 14 seconds of arc would be much too small. Aubert has also given values for black lines on a bright ground obtained in a similar way. He found that black lines subtending at the eye angles varying between 45 and 10 seconds of arc appeared too large by 34–44 seconds of arc. If Nelson's correction was applied in this case it would make the error larger and not smaller.

In most of the cases investigated by Bryan and Baker, it appears possible, from the values which they give, to estimate what the error will probably be for a particular individual with a particular type of test object; so that, this personal error being known, it should be possible to correct subsequent determinations. In other cases such correction would appear to be difficult if not impossible; for example, supposing the diameter of a highly-magnified object placed on the field of the microscope to be the subject of measurement by the spider-line micrometer. Then the edges of the image presented to the eye will appear to be surrounded by diffraction phenomena, which will vary with the working aperture of the optical system, and probably also with the colour of the light which is being used for illumination. Calculation of the distribution of intensity of light at the edge of the image will show that there is a gradient not unlike those met with in the case of absorption bands. The apparent edge, therefore, of an image with such a gradient of intensity of light will vary under similar conditions to those which have already been considered in the case of absorption bands. Suppose, for example, a black object on a bright ground to be under examination (*e.g.* a red-stained micro-organism), examined under direct blue green illumination, the image will appear to increase in size as the brightness of the illumination is diminished; or as the working aperture of the lens system of the microscope is diminished (the intensity being kept constant); or, lastly, as the eye of the observer becomes fatigued, with a corresponding rise in the value of his visual threshold. Under these conditions, determinations by the contact method will be found to give values which are not only inaccurate themselves, but which are also, it would seem, incapable of correction.

A practical point may be mentioned in this connexion. It

is possible that a more accurate estimate of the true size of the object might be obtained if measurements of the edge were made, firstly, of the object when seen as a black image on a bright ground, and then by a change in the method of illumination, as a bright object on a black ground. In the one case, according to Bryan and Baker's experiments, the measurements will be too large, and in the other case they will be too small by the amount which the eye subtracts from, or adds to, the true diameter, owing to the apparent edge of the gradient being situated at a certain distance away from the actual commencement of the gradient. An average of these two determinations might be found to give the actual diameter of the image without great error from physiological causes. This subject requires further experimental investigation.

Another type of error occurs in contact measurements, that is when the image of one object actually obstructs the image of the other object when the two are caused to overlap one another. For example, in the Abbé * apertometer it is necessary that the image of a cursor should be brought into contact with the edge of the restricting aperture of the objective lens system under measurement. Owing, however, to the fact that the image of the index lies in a plane below that occupied by the image of the aperture, the latter obstructs the former so that no apparent difference in the appearance to the eye results if the two are just mathematically in contact, or if there is a very considerable overlap. The eye, under these circumstances, has two alternatives before it—either to set the images so that it is seen that definite contact has just not occurred, or to set the images so that contact definitely has occurred, when, of course, there will be overlap by an extent which cannot be estimated by the eye. In either case there will be inaccuracy in the determinations. A method of avoiding this type of inaccuracy in the case of the apertometer has already been considered in a previous paper †. Briefly, the modification consists in such an alteration of the optical system that the indices lie between the restricting aperture under measurement and the eye of the observer. In the case of the apertometer this state of affairs is obtained by passing the illuminated beam in the reverse direction, *i. e.* in at the eyepiece of the microscope and out at the edge of the apertometer-plate to reach the eye of the observer.

* Abbé, Journ. Roy. Micr. Soc. 1880, p. 20.
† Hartridge, Journ. Roy. Micr. Soc. 1918, p. 337.

Section VI.

Measurements of Depth and Distance.

Three methods will be briefly referred to under this section :—

1. Monocular perception of depth by means of, *e. g.*, the micrometer microscope.
2. The monocular range-finder.
3. The binocular range-finder.

With regard to the first method, the usual procedure is to focus first one surface and then the other surface of the object whose depth is to be measured, and to ascertain the movement of the optical system that has been necessary in order to accomplish this. For example, in the case of a lens, lycopodium powder is sprinkled on the surfaces, and the grains on one surface and then the grains on the other surface are focussed in turn. This method of measurement is found to suffer from two errors, which are of a physiological nature :

(*a*) The range of accommodation of the eye.
(*b*) The depth of focus of the eye.

(a) *The range of accommodation of the eye.*—With regard to this, it will be seen that the accuracy of the determinations depends on the lower focal plane of the lens system preserving a constant relationship with the index of the scale of the instrument. But if a portion of this lens system, *e. g.* the crystalline lens of the eye, varies in its power, then the position of this plane must correspondingly alter. Suppose, for example, that the crystalline lens increases in power : then it can readily be shown that the lower focal plane rises with respect to the rest of the instrument, and *vice versa.* To eliminate this error it is necessary that there should be placed in the field of the eyepiece a suitable object on which the eye can be focussed, *e. g.* a cross-wire. If, then, the observer suitably adjusts the accommodation of his eye so that this cross-wire is always accurately in focus, to a very large extent errors due to change in the degree of accommodation of the eye will have been eliminated—but not entirely, owing to the second factor, which will now be discussed in some detail.

(b) *The depth of focus of the eye.*—Since, as has been mentioned above, the sensitive surface of the eye consists

essentially of a number of sensitive elements of approximately known diameter, arranged compactly together, it can be readily calculated that the eye will have a certain depth of focus which will vary with the aperture of the pupil, the smaller the pupil the greater being the apparent depth. Two imaginary planes may therefore be drawn in front of the observer and parallel to his retina. Images of external objects falling between these two planes will appear to be sharply focussed to his eye. Corresponding to each of these planes, there will be two planes situated on either side of the lower focal plane of the whole lens system of his eye, and these two planes will tend to separate further from one another as the diameter of the emergent pupil of the optical system of his eye is decreased in diameter. So long as the surfaces under measurement lie between these two planes, they will appear to form sharply-focussed images on his retina, so that unless these two planes are situated so closely together that their distance of separation is small compared with the accuracy to which the measurements of depth are required, error will be introduced into the measurements effected by the method. It will be seen that this error can be made smaller by causing the exit pupil of the eye and of any instrument in use to increase in diameter. In fact, it is probably correct to say that the largest possible aperture should be employed in the optical system that will enable sufficiently sharply-focussed images to be formed on the retina. It will be found, further, that the error decreases for a given aperture as the magnifying power is increased, and therefore the highest magnification should be employed which will permit the required measurements of depth to be made.

To return, for example, to the case of the determination of the thickness of a lens, mentioned above, that objective should be selected for use in the microscope whose working distance will just permit the two surfaces of the lens to be focussed in turn ; the objective should also have the largest aperture that can be obtained for this focal length ; and that magnifying power should be used in the eyepiece which will enable sharply-focussed images to be formed on the retina. This eyepiece should be fitted with cross-wires in order to eliminate the effect of accommodation of the eye, mentioned in the previous paragraph.

Monocular range-finders.—Since this instrument employs the method of coincidences, the principal errors to which it is liable have already been considered under Section III.

It should be pointed out, however, that there are two particular errors to which the instrument would appear to be liable. Firstly, that due to colour-parallax, when the objects under measurement reflect a considerable amount of coloured light; and, secondly, aperture-parallax, due to the exit pupils of the two optical systems—that to the right-hand side and that to the left-hand side of the base-line—not exactly corresponding to one another. It has been pointed out above that if the eye suffers from spherical aberration, this aperture-parallax may lead to a relative movement of the images in the two fields, which may introduce errors in measurement.

The Stereoscopic range-finder.—It would seem that colour-parallax is the only serious error to which this particular instrument is liable on physiological grounds. Its avoidance could, it would seem, be effected by either introducing a suitable colour-filter which would limit the rays reaching the eye to a nearly monochromatic bundle, or by introducing a pair of prism elements having dispersion without deviation which would cause the images formed on the fovea to overlap one another accurately, and would thus eliminate the effects of chromatic difference of magnification of the retinal images.

Section VII.

Comparison Methods of Colour and Intensity Measurement.

Two methods of measurement are largely used at the present day for a variety of purposes:

(*a*) Flicker photometry.
(*b*) Comparison photometry.

With regard to the former, physiological variations have already been very thoroughly studied. With regard to the latter, it would seem that the conditions necessary for accuracy are not so well known. In colorimeters, polarimeters, spectro-photometers, and the like, three types of field are to be met with:

1. The circular field divided into two parts by a diagonal.
2. The circular field containing a smaller, centrally-placed subsidiary field (the so-called bull's-eye type of field).
3. The field divided into three parts by two parallel lines, so that the central strip forms one field and the two side strips form the other.

The errors to which measurements made with instruments which possess these three types of field are liable are, firstly, retinal after-images, and, secondly, retinal fatigue images. It is found that in all three types of field the attention appears to be concentrated on one particular part of the field, so that a definite pattern is always presented on the same parts of the retina. Supposing that at the commencement of measurement, one part of a field is always the brighter, and that adjustment be made by causing the two fields to become more and more alike until equality is reached, then that part of the retina corresponding to the brighter field will suffer to some extent from fatigue, and will also be occupied by what is known to physiologists as a negative after-image. Both these factors operate together to reduce, apparently, the sensitiveness of the retina, so that a brighter image has to be presented to this part of the retina than that required to stimulate the other portions on which the brighter image did not at first fall. The result will be that, when equality appears to have been obtained, the image which was initially the brighter is to some extent still the brighter, although it appears equal to the less bright part.

The bull's-eye type of field has, in addition to this error, a special error of its own. It is well known physiologically that the portion of the retina which corresponds approximately to the optical axis of the eye is the fovea. It is this part on which images are caused to fall by directing the eye towards them in order that the highest definition of visual acuity should be obtained. The fovea differs histologically from the more peripheral parts of the retina in having cones only, whereas in the more peripheral parts of the retina, rod retinal elements tend to predominate. It is well known that whereas the cones are used for the appreciation of colour, the rods are colour-blind, and therefore in the bull's-eye type of field the central area falls on the more colour-sensitive part and the annular field on the less colour-sensitive part of the retina. If, then, comparison of colours is being made, it will be readily seen that the bull's-eye type of field is specially objectionable. But, further, even if it is only intensity comparisons which are being made, the cones are known to function principally in day-time and the rods principally in night-time, and therefore, according as the general intensity of illumination of the fields is high or low, the centre or the periphery will appear to be relatively the brighter. In instruments allied to the spectro-photometer, a fourth type of field is occasionally met with *, the eye focussing sharply

* Hufner, *cf.* Milne, Proc. Opt. Convention, 1905, p. 178.

a narrow vertical slit in the plane of which a spectrum is produced. This spectrum is divided into two contiguous portions, one of which has passed through one limb of the instrument and the other through the other limb. Suitable adjustment is made so that these two halves appear to be alike in brightness, or, in certain cases, alike in colour. The disadvantages of this narrow field are that the line of junction of the two fields which are to be compared is so small compared with their total area, and it is usually found that the line of junction cannot be increased by making the slit wider without at the same time introducing an obvious difference of colour on the two sides of the field seen by the eye, which would, in its turn, have the effect of making comparison difficult.

The design and construction of colorimeters and spectrophotometers with fields which do not suffer from the errors that have been pointed out above, have been described in previous papers*. The alternative type of field embodied in these instruments consists of a number of narrow strips having a very sharp line of separation between them, each alternate strip being illuminated by light from one or other limb of the instrument. Adjustment is made until the strips, regarded as a whole, appear to become alike in colour or intensity, so that their contours disappear, and the field appears to become uniform without any lines of separation.

With regard to measurements involving the determinations of shades of colour, it has been shown by Steindler and others that the sensitiveness of the eye to change of wavelength varies with the part of the spectrum under investigation, being greatest in the yellow-green and least in the red. Where mixed light, *e. g.* daylight, is concerned the eye appears to be most sensitive to variations about the neutral point. For example, the addition of a weak coloured light is more readily detected if it is added to a white light than it it be added to a coloured light of the same intensity as the white. This effect can, in practice, be frequently obtained by the use of colour-filters of suitable strength and of complementary colour to those under estimation. For example, a suitable piece of signal-green glass will be found to facilitate the change in the tint which occurs when vermilion oxy-hæmoglobin changes to crimson reduced hæmoglobin. Through the signal-green glass the oxy-hæmoglobin appears orange, changing through neutral-grey to violet as it becomes reduced.

* Hartridge, Proc. Camb. Philo. Soc. xix. p. 271 (1919); Hartridge, Journ. Physiol. l. p. 101 (1915).

Summary.

1. The limits to the accuracy of observation and measurement considered in this paper are those introduced by the imperfections of the eye.

2. Experiment shows that even under ideal conditions the centres of two lines or dots must subtend at the eye an angle of approximately 60 seconds of arc in order that they shall be seen as two. When the conditions are not ideal a much greater separation is required, even by normal-sighted people. People with defective vision require correspondingly greater separation.

3. Formulæ have been published giving the limiting useful magnifying power for optical instruments of different aperture. Thus, in the case of the microscope, Helmholtz has given 266 × N.A. as the limit; while Abbé's formula is 1000 × N.A. Both are too low, even for persons with acute sight, and should be replaced by 2000 × N.A. ÷ Snellen's standard of visual acuity for the observer, this standard having been determined previously by experiment.

4. The visual acuity of the eye for the positions and movements of contours is nearly ten times greater than it is for the resolution of double points, and lines. This fact is made use of in mensuration.

5. The method of coincidences is found by experiment to give very accurate results. It has been found that the error at the eye is less than 10 seconds of arc ($0{\cdot}76\,\mu$ error in the setting of the image on the retina).

6. Experiments by the author show that blurred lines, *e. g.*, absorption bands, are set into coincidence with approximately the same accuracy as are sharp lines. In this connexion it should be noted that, owing to diffraction and chromatic aberration in the eye, the images of sharp lines are diffuse, so that the retinal images in the case of blurred and sharp lines are not very different.

7. The coincidence method of measurement is found to be liable to a number of errors due to the physiology of the eye; these are:—(*a*) line thickness, (*b*) malfocus, (*c*) colour, (*d*) intensity of light, (*e*) visual threshold, (*f*) non-coincidence of exit pupils. The means of avoiding these errors is described.

8. The interpolation method of measurement used, *e. g.* in the chemical burette and the slide-rule, is much less accurate than the coincidence method. Otherwise the errors to which the method is liable are similar to those mentioned above in the case of the coincidence method.

9. The contact method of measurement used in the spider-line micrometer eyepiece is liable to error due (*a*) to irradiation, (*b*) to contrast. Bryan and Baker's experiments showed that the sizes of black objects are overestimated, while the sizes of white objects are underestimated. On the other hand, Nelson has given rules for the correction of the measurements made on objects under different methods of illumination: it is found, however, that when the correction is applied it tends to make the error larger. Further work is required to elucidate this problem.

10. Measurements of depth and distance are, in the case of monocular methods, liable to two errors: (*a*) due to the accommodation of the eye, (*b*) due to the depth of focus of the eye. The means of avoiding these errors is described. The errors to which monocular and binocular range-finders are liable are considered, and deductions drawn as to the relative accuracy of the two instruments.

11. The methods used in measuring colour and intensity are described. The form of the "fields" used in such instruments is criticized. An alternative type of field is suggested by which more accurate determinations may be made. For colorimetric work the use of complementary illumination is suggested.

V. *On Images obtained by means of a Semi-infinite Obstacle.*

By SATYENDRA RAY, *M.Sc.* (*Allahabad*) *.

[Plate III.]

THE object of the present paper is to point out that if light is passed through a number of parallel slits, which therefore serve as independent sources, a straight edge held parallel to the lines will produce on a screen an image of the slits. The image is inverted, as can be easily seen by stopping any particular slit with a thin rod.

The image is best described as being very similar to the diffraction image that would be expected by placing a slit just outside the straight edge. This can be demonstrated by bringing another edge from the opposite side. In fact a little attention will show that the second edge carries with it its own image of the pattern, and that the slit forms the clearest image when the images separately due to the two edges suitably overlap. The above phenomenon is the analogue, at any rate for linear objects, of the effects

* Communicated by Prof. A. W. Porter, F.R.S.

produced by a pin-hole camera, or again, of the images obtained by means of an opaque disk shown by Prof. A. W. Porter *.

Variation of the distances of the object and the screen makes the size of the image vary according to the pin-hole camera law. If the light is passed through Wratten filters no perceptible change in the image is noticed. The image is produced by a cylindrical edge equally well, a brass tube or the wall of a glass bottle serving the purpose of an edge.

We can explain the phenomenon easily if we take into account only the first maximum of the diffraction fringes due to each source and neglect all the rest. That we are justified in doing this is seen by taking a photograph of the image produced by a single slit source. In this case the first diffraction band is by far the most prominent, as can be seen in fig. 1 (Pl. III.), which gives us the fringes corresponding to a thin slit source less than 1/5 mm. in width. As the position of the different maxima from the edge of the geometrical shadow due to a slit is given by

$$x=\sqrt{\frac{q(p+q)}{p}(2n+1)\lambda},$$

where p and q are distances of the slit and screen from the edge, and as the image follows the pin-hole camera law

$$x=a\frac{q}{p},$$

maxima other than the first for each slit can only tend to make the pin-hole camera effect confused.

In figs. 2 and 3 (Pl. III.), we have the photographs of the images formed by a razor blade held parallel to the wires of a wire grating at a distance of 50 cm. from it, the distances of the screen from the edge being 100 and 200 cm. respectively. The four wires and the five slits of the grating may be seen in the image.

The writer stumbled upon the experiment accidently at Meerut College, India in 1911, by sunlight, passing through a split bamboo screen, forming such an image by the edge of a book. The photographs reproduced here were taken at University College, London. I am grateful to Prof. A. W. Porter for his kind interest in the experiment and for guidance.

University College, London.
24th Nov., 1922.

* Phil. Mag. xxvii. p. 673 (1914).

Fig. 1.

Fig. 2.

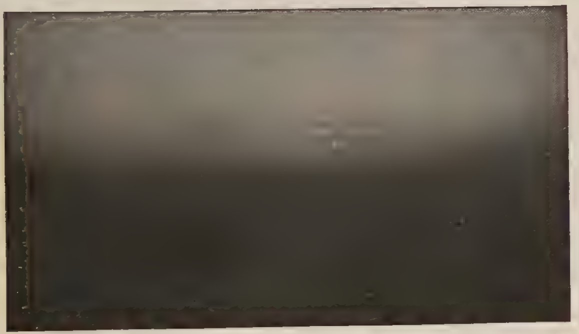
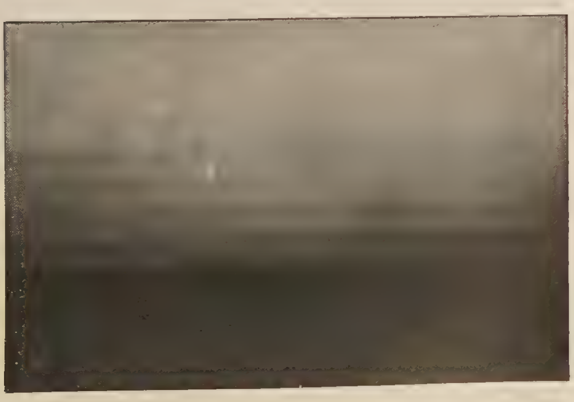


Fig. 3.



VI. *The Recording Ultramicrometer: its Principles and Application.* By John J. Dowling, *M.A., F.Inst.P., M.R.I.A.**

In a short paper before the Royal Dublin Society †, I gave a preliminary account of the employment of an oscillating valve circuit as an ultramicrometer. I now propose giving a fuller account of the principles of the device, together with particulars of some further tests of the apparatus ‡.

The ultramicrometer was a natural development of an electric valve method I described a short time previously § for measuring small-capacity variations. This latter method required an external source of high-frequency alternating current, and it occurred to me to try to incorporate in the apparatus itself the necessary elements of an oscillation circuit so as to make it self-contained. It will, of course, be understood that an apparatus which records small capacity variations will be immediately suitable for measuring minute movements, inasmuch as the capacity of a condenser varies with the separation of the plates.

A principle of a completely different kind was employed in the ultramicrometer device described by Whiddington ‖. This was an application of the well-known "heterodyne" method first described a year earlier by American workers¶, The characteristic "heterodyne note," due to the beating of the electric oscillations in two oscillating circuits nearly in tune, varies if the capacity of one of these circuits is altered slightly. English writers have occasionally confused Whiddington's apparatus with that here dealt with. There is nothing in common between them beyond the fact that radio apparatus is employed in both. The complexity of Whiddington's heterodyne method, due to the employment of *at least* two oscillating circuits; the fact that only acoustic effects are observed, which, in practice, involves some strain

* Communicated by the Author.

† Proc. Roy. Dubl. Soc. vol. xvi. p. 18 (March 1921).

‡ Other accounts were published in 'Engineering,' Sept. 8th, 1921; 'Scientific American,' Feb. 1922; *Die Umschau*, March 1923; besides many summaries of these. The apparatus was exhibited at the British Association at its Edinburgh meeting, 1921.

§ Proc. Roy. Dubl. Soc. vol. xvi. p. 17 (Feb. 1921).

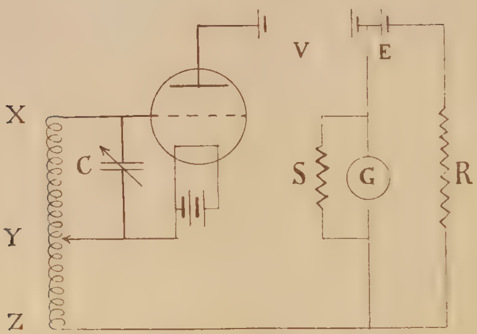
‖ Whiddington, Phil. Mag. vol. xl. (Nov. 1921).

¶ Hyslop & Carman, American Phys. Soc. Dec. 1912; Phys. Rev. vol. xv. p. 243 (1920).

on the observer *; and the fact that the mean separation between the condenser plates (a small and difficultly measurable quantity) has to be determined—all render the "heterodyne" method unsatisfactory. I should like to testify, however, to the ingenuity which Whiddington has shown in the application of the method.

A simple form of oscillating circuit is shown in fig. 1, in which, for the present, we shall imagine S, G, R, and E to be replaced by a milliammeter. The condenser C is of the variable sector type, and the coil XYZ consists of about 150 turns of bare copper wire wound in a helix on a wooden frame of about 6 in. square section and 2 feet (60 cm.) long.

Fig. 1.



The valve may be an ordinary "R-type" valve, with a 4-volt filament battery, and the anode battery V is of 50 volts. For certain positions of the filament connexion Y on the coil, the circuit will oscillate, and it is found that the milliammeter in the anode circuit then shows an increased reading. This effect is much greater when the grid potential is such that the valve is operating near the lower (or upper) bend of its characteristic before oscillations begin. For too small, or too large, values of the capacity the oscillations usually cease, and between these values the milliammeter readings vary with the capacity.

The curves in fig. 2 show some observations originally taken in order to get some idea of the conditions obtaining. The condenser then available was of limited range (100 to 1200 $\mu\mu$ farads), and it was not possible to trace the complete range of oscillation for any one fixed position of Y. A series of curves are therefore given, the number on each being the

* *Cf.* Carman & Lorance, Phys. Rev. vol. xx. p. 715 (Dec. 1922).

length in cms. in the part XY included in the oscillation circuit ("tuned grid circuit"). A detailed discussion of these curves is, however, beside our purpose. It is sufficient to notice that many of them are quite steep in gradient, and, what is much more important, several (such as those for 40 cm. and 44 cm.) have a pronounced hyperbolic form over a good part of the range shown. In many places we find a variation of 1 milliampere per 100 micro-microfarads, or, roughly, 10 *micro-amperes per centimetre variation of capacity.*

Let us, now, consider a condenser formed of two circular disks of radius r cm. and x cm. apart. In electrostatic units,

$$C = \frac{r^2}{4x};$$

hence

$$\frac{dC}{dx} = -\frac{C}{x}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Assume $r=5$ cm., $x=0{\cdot}025$ cm.; so that $C=250$ cm. and $\frac{dC}{dx}=-10{,}000$. Thus a decrease of 1 cm. in the capacity accompanies a separation of the plates of only 1/10,000 cm.

At the point 277 $\mu\mu$ f. (*i. e.* 250 cm. capacity e.s.u.) on the curve 40 (fig. 2) the slope is roughly 8 micro-amperes per cm. capacity. Therefore, if the condenser just considered were employed in this case, we should expect to obtain an anode current *variation* of 8 micro-amperes for a relative displacement of the condenser plates of 1/10,000 cm.

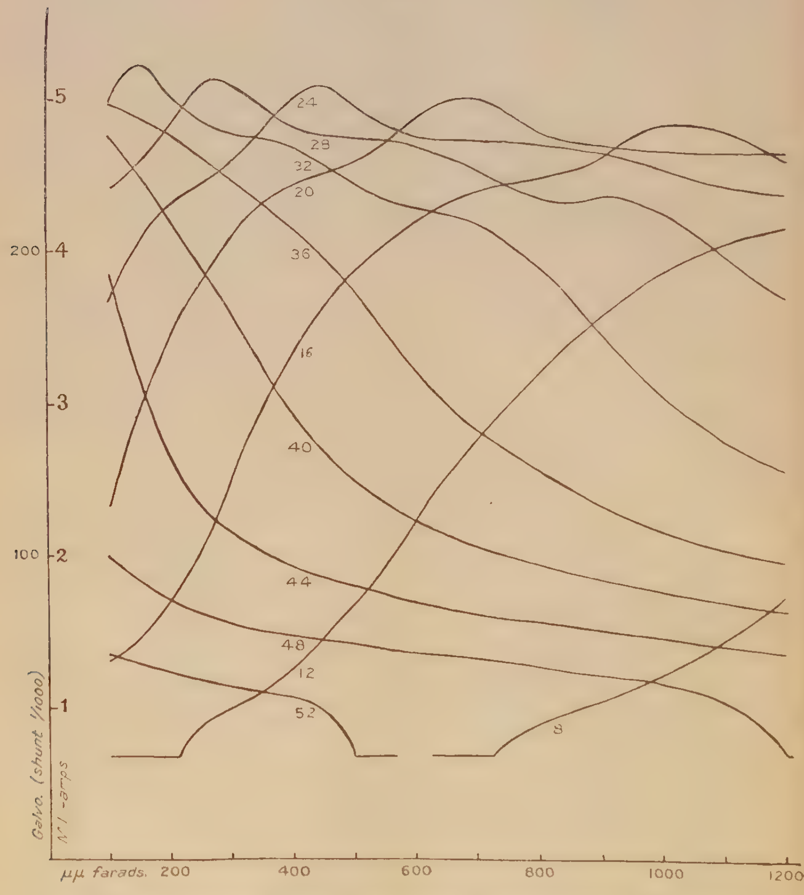
It must however be noticed that the plate current, which undergoes this variation, is already of a strength of over 3 milliamperes.

The milliammeter, which we have so far supposed to have been used to measure the anode current, is incapable of registering this small variation, and a special device is substituted, as shown in fig. 1, to enable this to be done.

A resistance R and an extra battery E are connected in series with the anode battery V, and the value of R is adjusted so that the anode current through it produces a drop of potential equal to the P.D. maintained by E. There will then, clearly, be no difference of potential between the points to which the galvanometer G is shown connected. If, however, a change takes place in the anode current, this no longer obtains, and the *excess* (or *defect*) current divides between G and RE according to the laws of parallel

circuits. In other terms, the galvanometer is completely shunted for a certain "zero" value of the current, but not

Fig. 2.



for other values: hence the name "zero shunt," by which I call the device. The theory is as follows:—

Let i be the "zero" anode current for which we desire the galvanometer to show no deflexion, g the resistance of the galvanometer circuit, R and E the resistance and e.m.f. constituting the "zero shunt." Then

$$iR = E. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Suppose i increase to $i+\Delta$, and let a, b be the currents

through the galvanometer circuit and R respectively :

$$ag = \mathrm{R}b - \mathrm{E}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

$$i + \Delta = a + b. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

Thus $\qquad \mathrm{R}(i+\Delta) = \mathrm{R}a + \mathrm{R}b = \mathrm{R}a + ag + \mathrm{E}\,;$

$$\therefore \qquad \mathrm{R}\Delta = (\mathrm{R}+g)a.$$

Hence

$$a = \frac{\mathrm{R}}{\mathrm{R}+g} \,.\, \Delta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

which expresses the fact stated above. It will be clear that R and E can have any values satisfying (2), but if R is made large compared with g, practically the *whole* variation Δ of the anode current passes through the galvanometer.

Returning to the consideration of the "capacity current" curves, it will now be evident that the complete arrangement shown in fig. 1 may be extraordinarily sensitive to changes in capacity if a moderately sensitive galvanometer is employed. Suppose the galvanometer used gives a deflexion of 100 scale-divisions per micro-ampere—an ordinary laboratory value. Reverting to our previous numerical example, 1/10,000 cm. displacement of the condenser plate caused an anode current variation ("Δ") of 8 micro-amperes. With a high-resistance zero shunt the galvanometer in question should show 800 scale-divisions ; that is to say, *one* scale-division represents a plate movement of 1/8,000,000 cm. This enormous sensitivity might seemingly be exceeded by selecting curves of steeper gradient : but in doing so there are other conditions which, in practice, it is very important to keep in view. By deriving an expression for the sensitivity of the apparatus we shall arrive at these.

Let us define the micrometric sensitivity (M) of the apparatus as the displacement of the condenser plates necessary to produce a deflexion of one galvanometer scale-division. Let

a = galvanometer current (amperes),

G = galvanometer constant = a/D,

D = galvanometer deflexion due to current a,

$f = \dfrac{a}{\Delta} = \dfrac{\mathrm{R}}{\mathrm{R}+g}$ (as above, equation (5)),

$\Delta = di$, where i = anode current,

s = slope of "current-capacity" curve (fig. 2) $= \pm \dfrac{di}{d\mathrm{C}}$,

C = (variable) capacity in e.s.u. = $\mathrm{A}/4\pi x$,

x = separation of condenser plates.

Then

$$G\frac{dD}{dx}=\frac{da}{dx}=f.\frac{di}{dx}=f.\frac{di}{dC}.\frac{dC}{dx}=-\frac{f.s.A}{4\pi x^2}=-\frac{f.s.C}{x}. \quad . \quad . \quad . \quad (6)$$

Hence

$$M=\frac{dx}{dD}=-\frac{Gx}{f.s.C}=-\frac{4\pi G}{f.s.A}x^2. \quad . \quad . \quad . \quad (7)$$

An important deduction may now be made. If the sensitivity remains at a constant value S over an extended range of displacement of the condenser plates,

$$M = \text{constant} = -\frac{4\pi G}{fsA}x^2 = -\frac{AG}{4\pi f}\cdot\frac{1}{C^2 s}.$$

Thus

$$\pm C^2\frac{di}{dC} = C^2 s = -\frac{GA}{4\pi f M} = \text{constant}. \quad . \quad . \quad . \quad . \quad (8)$$

Integrating,

$$\frac{GA}{4\pi f M}\cdot\frac{1}{C} = j \pm i, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where j is a constant, and either the + or − sign may be taken.

Both these forms represent hyperbolas, and I have already pointed out that many of the "Ci" curves of fig. 2 approximate over large parts of their contours to this type. This is especially so on the parts where s is negative, as, for example, the greater part of curve 40, including the region chosen for our previous calculation. By adjusting the apparatus to function in such a region, a wide range of practically constant sensitivity is available. This facilitates very considerably the practical application of the micrometer device. For, by working within this range, it is not necessary to know the actual *total* separation (x) of the condenser plates; and indeed one can generally judge the right distance by the eye. Furthermore, the actual calibration of the apparatus may be simply done by giving one of the plates a predetermined minute displacement, and we are justified in assuming that the value of s so obtained is still applicable, even though the plates are subsequently moved very far from the point of calibration (provided that the capacity remains within the "hyperbolic" range).

Assuming the apparatus working within this "hyperbolic" range, we then have $C^2 s$ constant (eq. 8) and (7) the sensitivity

$$M=\frac{Gx}{fsC}\propto\frac{GxC}{j}\propto\frac{G.A}{f} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The proportionalities refer to a particular chosen *coil* adjustment, but the equality on the left is applicable generally. With a given adjustment of the coils, therefore, the sensitivity is *proportional to the area* of the condenser plates employed and to the sensitivity of the galvanometer. The factor f is in practice nearly unity. From our definitions of M and G a great sensitivity means a very small numerical value, so that (10) indicates that higher sensitivity is obtainable with condenser plates of small diameter very close together.

Hitherto we have considered the "current-capacity" curves from the point of view regarding the suitability of the apparatus as a micrometer device. On some such curves there are sections where the slope is sensibly uniform—*i. e.* where a *linear* relation exists between the anode current and the capacity. Professor Carman *, of Illinois, has made use of this feature of the ultramicrometer circuit to measure minute capacity differences in connexion with his work on the dielectric constant of gases.

Besides these two "useful" portions, many of the curves show humps and shoulders, and frequently two distinct maxima. Some experiments have been carried out to determine, if not the origin of these irregularities, at least how best to modify the circuits so as to reduce or eliminate them. Incidentally, some experimental confirmation of the theory of the functioning of the circuit has been forthcoming. We now proceed to deal briefly with these questions.

So far the apparatus under consideration has been a simple "tuned grid" circuit with a rather large coil. I have found that much better results are obtainable with this same circuit, if a shorter coil of rather wider diameter is employed (20 cm. long; 20 cm. diameter: 42 turns No. 22 copper). With about 15 turns in oscillating grid circuit, a very large range of high uniform sensitivity is obtained.

However, the single coil, if it is to permit of "coupling" adjustment by alteration of the connexion Y, has to be wound with bare wire and the turns spaced. Naturally it is not compact. A much more satisfactory arrangement is obtained by using two "pancake" or "basket" coils. These can be mounted so that the coupling is adjustable by transverse or axial displacement of one of the coils.

* Carman, *loc. cit.*

I have generally employed a pair of coils of 10 cm. diameter, wound with about 150 turns No. 28 or No. 36 D.S.C. or enamelled copper wire. A variometer can also be used. There are, in all, four simple circuits which serve almost equally well when properly set up and adjusted :—The "tuned grid" (fig. 1), already dealt with : "tuned plate" ; "Hartley" (fig. 3) ; "modified Meissner" (fig. 4). The last-named is not so simple as the others, requiring a third

Fig. 3.

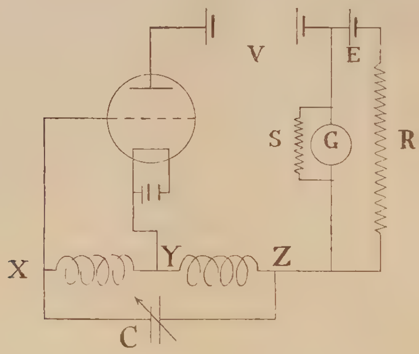
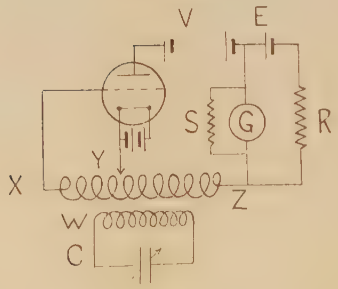


Fig. 4.



separate coil for the oscillation circuit. To complete the consideration of practical circuits, we will give some results obtained with circuits made up with the flat coils.

In order to display the extent of the range of constant sensitivity, a special "micrometer" condenser has been employed in some cases. This consists of the graduated fine adjustment movement of a Leitz microscope, to the table of which one of the condenser plates is fixed, while the other is carried at the end of a glass rod projecting from

the microscope tube. The micrometer screw is graduated to read 1/200 mm.

Fig. 5 shows the micrometer curves for a "tuned-grid" circuit, the coils of which were mounted parallel and about 1 cm. apart. The successive curves correspond to different degrees of coupling obtained by lateral displacement of the coils. It is clear that the desired linear form is possessed over very wide ranges at certain adjustments.

Fig. 5.

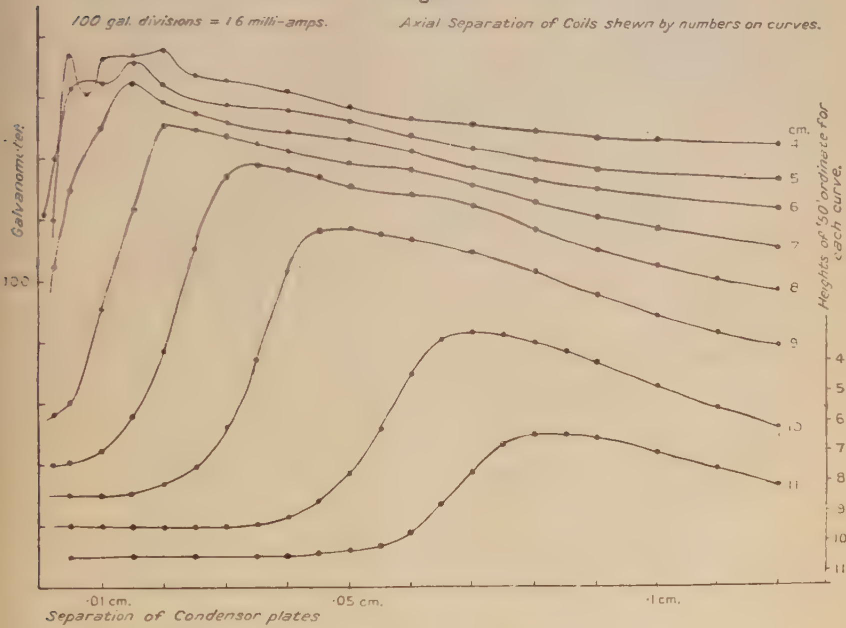
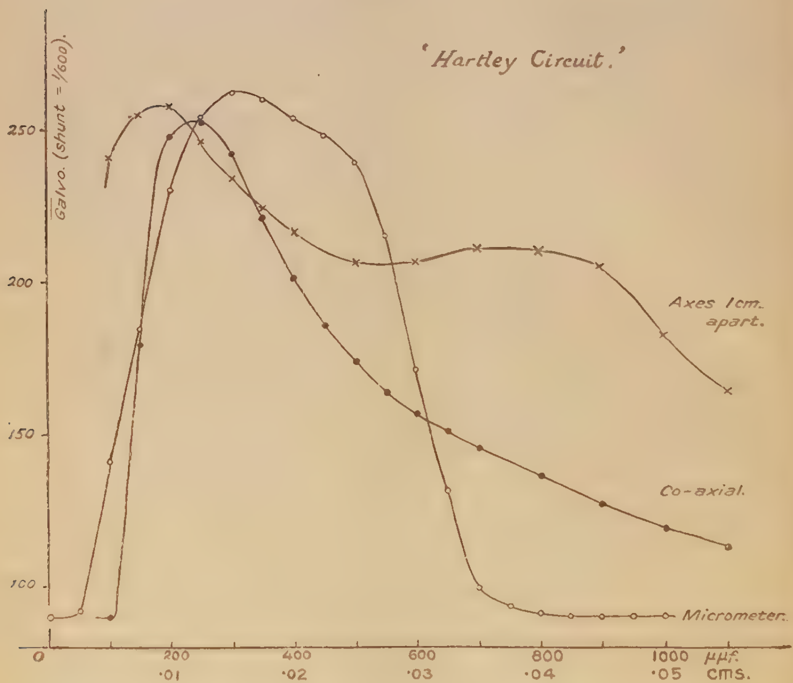


Fig. 6 shows three curves for a "Hartley" circuit constructed from the same pair of coils, mounted in series opposition (fig 3). We may note that the adjustment is quite critical, a displacement of 1 cm. from the co-axial position distorting the "capacity-current" curve completely. The "micrometer" curve is taken in the co-axial position, and shows an almost perfect linear form on either branch, great sensitivity being attainable. The adjustment is critical also in respect to closer approach of the coils. A Hartley circuit usually ceases to oscillate if we mount the coils too near together; in fact, the best "micrometer" adjustment is just on the border-line.

The functioning of these circuits is often found to be

improved by the introduction of a "blocking" condenser (say ·05 or ·1 μ f.) between the anode and the point Z. A large condenser like this has two beneficial effects. In the first place, it minimises external "capacity" effects. The zero shunt, galvanometer, and H.T. battery are of necessity external to the rest of the apparatus, and the first-mentioned has occasionally to be handled by the observer. The presence of a large capacity in this part

Fig. 6.



of the apparatus tends to swamp transient capacity effects due to movements of the observer near by. In the second place, the secondary maxima and shoulders on the curves, if present, are considerably diminished or entirely eliminated in the presence of the blocking condenser. I have to thank Miss Katharine Preston, who investigated this latter point for me. She found that by starting with a small blocking condenser and gradually increasing it, one of the maxima moved, as it were, across the curve and ultimately disappeared with a sufficiently large capacity.

Miss Preston also investigated the effect of resistance in these oscillating circuits *. As one would expect, this is very considerable. Although, in general, it may be stated that for high sensitivity (steep curves) the aim should be to keep the resistance in the oscillating coils low, nevertheless it is possible to overshoot the mark in this respect.

With coils of low resistance the effect of change of capacity is so great that instability sets in. A sort of hysteresis is displayed. The circuit refuses to start oscillating until the capacity is altered to a value well within the oscillating range; while, when oscillating, the circuit continues to do so for values of capacity for which it previously refused to respond. In each case oscillation starts or stops suddenly. A clue to this behaviour is forthcoming when we consider the probable theory of the micrometer circuits.

The usual methods of formulating the necessary conditions for oscillation maintenance in a circuit are, of necessity, only approximate. They do not yield a value for the change in the d.c. component of the anode current. This is, however, considerable, except in the case when the valve is oscillating about the mid-point of its characteristic. It is this anode-current effect that we employ in a micrometer circuit. In the absence of a satisfactory theory a general statement of the case is all that is possible.

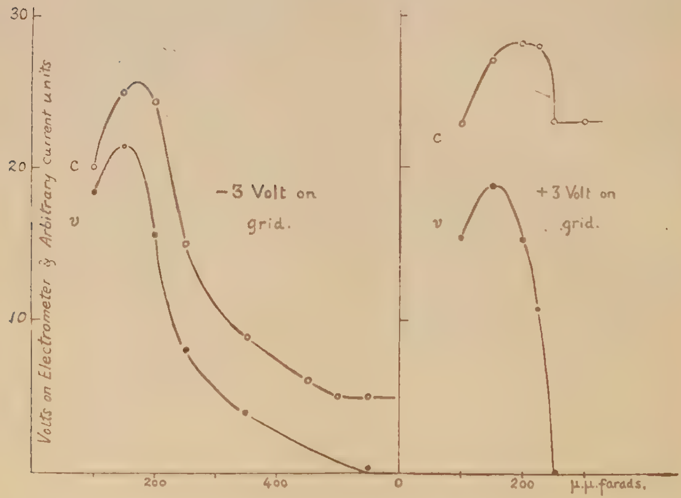
If the "tuned circuit" has *little ohmic* resistance, it acts as if it possessed very *great* resistance at its natural frequency. Consequently the capacity adjustment must be carried far into the range of oscillation before this sets in. When oscillations have started, however, the effective resistance of the valve will drop. This, to some extent, compensates the reactance effect. The reverse capacity adjustment can then be carried out without stopping the oscillations. This, rather rough, version of the matter serves to explain the hysteresis effect. The result of inserting, say, 500 or 1000 ohms in series with the condenser is to *reduce* the "resonance" reactance effect. The oscillations now start and cease at almost the same value of the capacity, and no noticeable "hysteresis" is displayed. It is an attractive problem to consider whether the ideas here put forward have any bearing on wireless telegraphic practice; but this is beside our present interest.

* Dowling & Preston, Phil. Mag. vol. xliii. p. 537 (March 1922).

The other effect of resistance is to reduce the amplitude of the oscillation in the tuned circuit. This is accompanied by a corresponding reduction in the anode current.

The only other factor to be considered is the grid potential. In this connexion it will be observed that the circuits described have a direct metallic connexion between the grid and the negative terminal of the filament. With some valves a few volts—usually negative—have to be applied to the grid. I have not tried the grid-leak-condenser device for this purpose, as the necessary high resistance for the purpose would either be too bulky or else unsteady.

Fig. 7.



Some experiments were carried out for me by Mr. A. Fynn, S.J., to investigate experimentally how the amplitude of the oscillation in the tuned circuit depended on the grid potential; how it varied with the capacity in circuit; and, incidentally, the relation between it and the anode current. The usual ultramicrometer circuit was set up, using a variable air condenser, and a sensitive Dolezalek electrometer, connected ideostatically, was used to determine the r.m.s. potential on the condenser. Fig. 7 shows the type of results so obtained. For moderate negative values of the grid potential (left-hand curve) we obtain considerable anode-current variation, and the amplitude of oscillation

rises and falls similarly with the capacity changes. There is a noticeable lag of the anode current behind the oscillation voltage. For the other curve, taken with a positive grid potential which corresponds to a point near the middle of the valve characteristic, the result is very different. Little anode-current variation occurs, while the strength of the oscillation is also somewhat less. This is quite what one would have expected. The second case is that which is best for the production of pure harmonic oscillations; the first is, of course, the best for micrometer applications. (This is also the more efficient adjustment of the valve as a generator of oscillations, although there are generally harmonics present.)

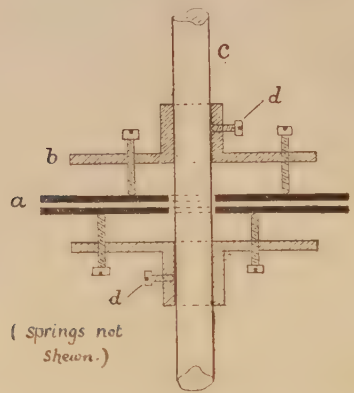
One point remains before we conclude with the general principles of these circuits. So far the coils and condensers considered have been such as to generate radio-frequency oscillations (10^5 to 10^7 ∼). I have, however, been able to obtain similar behaviour from circuits oscillating with frequencies as low as 500 ∼. These require large capacities naturally and cannot be employed as micrometer devices, but capacity measurements at audio frequencies could be carried out with them.

We now proceed to give an account of some of the results obtained with these circuits and to discuss certain of their applications. It may be mentioned in this connexion that in most cases my aim has been to determine the best results obtainable under ordinary working conditions. None of my experiments have yet been carried out with the apparatus completely shielded. This could easily be done with the more recent forms of the apparatus, in which the galvanometer, micrometer circuit, and batteries altogether occupy about two cubic feet of space. However, my present experiments do not call for the highest sensitivities, and I dispense with this refinement. Even without such precautions the apparatus is extraordinarily steady, due no doubt to the damping and inertia of the galvanometer.

To measure moderate sensitivities the micrometer condenser already referred to is very convenient, but for very high sensitivities I use the condenser arrangement shown in in fig. 8. The two plates (*a*) are supported on the ebonite collars (*b*) by suitable adjusting screws and springs. The ebonite collars are each clamped to the steel rod (*c*) by three set screws (*d*) which grasp the rod at points 5 cm. apart. The condenser plates are circular steel disks, 10 cm.

diameter, turned to a plane surface and bored 3 mm. larger than the steel rod, which is of 12 mm. diameter. The rod stands vertically and carries a small wooden platform above. To move the plates together, weights are placed on this platform which obviously causes the plates to approach by the amount of compression produced in the 5 cm. length of the rod by the added weight. Slight bending, due to

Fig. 8.



excentric location of the weight, will produce little error in view of the construction. The following table gives the results of one such determination :—

Table I.

Compression weight (kilogrammes).	Galvanometer scale reading.	Deflexion for 500 grammes.
0·0	−187	—
0·5	−105	82
1·0	− 27	78
1·5	+ 45	72
2·0	+121	76

The differences between the numbers in the last column are partly due to too tight a fit of the ebonite collars of the condenser on the rod, and partly to want of screening of the apparatus. The observations indicate an average deflexion of 154 divisions per kilogramme. The calculated compression corresponding to this is $2{\cdot}16 \times 10^{-6}$ cm. One galvanometer division should correspond to $1{\cdot}4 \times 10^{-8}$ cm.

This would probably have been the sensitivity under "screened" conditions, but the actual observations cannot be regarded as showing this sensitivity. The maximum variation of the numbers in the last column of Table I. from their mean is 5. Thus there would be an uncertainty in any reading of $5 \times 1{\cdot}4 \times 10^{-8}$ $(=7 \times 10^{-8})$ cm.: that is, a sensitivity of 10^{-7} cm. per galvanometer scale-division can be reckoned on.

The application of the recording ultramicrometer to the determination of minute displacements, changes of length, and the like, does not call for a detailed description. In view of the wide range of constant sensitivity, no very delicate adjustments need be provided for the condenser plates. In most cases, I mount one on a micrometer screw for calibration purposes, and the other with three adjustment screws like those shown in fig. 8. For very fine measurements the micrometer screw is made to work against a weak spring, which is attached to the much more rigid cantilever plate-support: this well-known device enables a minute plate movement to be effected, but requires preliminary calibration in some other way. The members supporting the plates are of course attached suitably to the two systems whose relative movement is to be recorded.

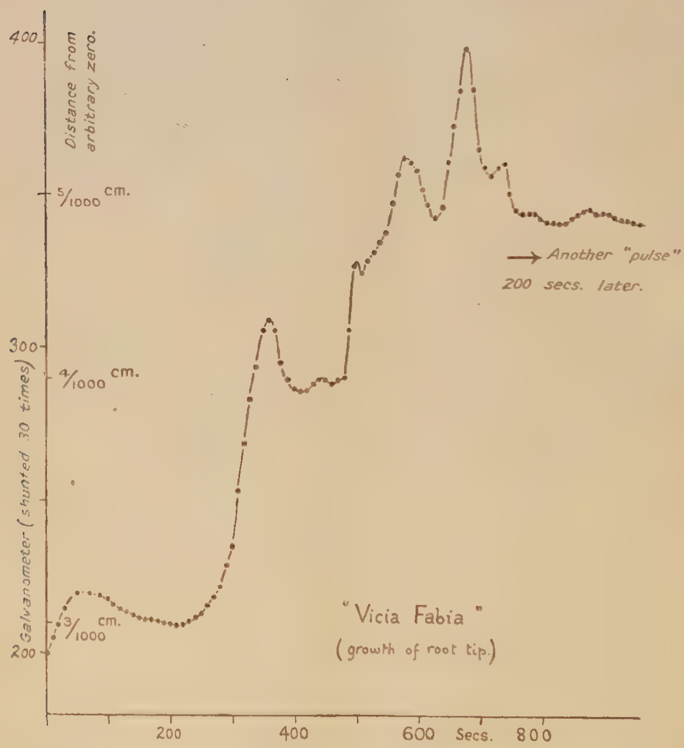
An important application of the device is in the construction of an extensimeter. Ewing, writing in the 'Encyclopædia Brittanica,' states that, using the customary 8-inch specimen, "it is desirable to read the extension to, say, 1/50,000 inch, if the modulus of elasticity is to be found with fair accuracy, or if the limits of proportionality of stress and strain are under examination." In other words, it is desirable to measure to within two-millionths of a centimetre per centimetre length. Our apparatus could easily record this latter amount, and *it would not be necessary to use a specimen even one inch long*. Furthermore, by attaching condenser plates so as to record the *lateral* strain, the determination of Poisson's ratio (and therefore all the elastic constants) would be possible with quite small specimens.

It may be objected that there is little space available around the specimen in a testing machine. In applying the method the condenser plates, themselves quite small, are the only parts that have to be near the specimen. The oscillation circuit proper is situated several feet away, and connected to the condenser by a pair of wires about an inch apart (*separated, not twisted*). I have tried out a device of this kind and found it answered satisfactorily. There may

be slight "capacity" effects if the operator approaches too near the condenser wires, but it is easy to arrange the apparatus so that these can be screened.

A very important consideration in a recording device is whether it responds to rapid variations of the observed phenomenon. With the usual ultramicrometer circuits the only limit in this connexion arises in the galvanometer. An excellent example of the utility of the ultramicrometer for following minute irregular movement arises in the study of plant growth *. Using a galvanometer, rated

Fig. 9.



as having a period of 2·5 secs. but almost dead-beat, it was apparently possible to follow movements whose duration was about 10 seconds. Fig. 9 shows a series of observations of the galvanometer when the root-tip of *Vicia Fabia* was under examination. The lower condenser plate

* Dowling, 'Nature,' June 23rd, 1921.

was mounted on a micrometer screw; the upper, almost parallel to it, was carried on the end of a light flat spring. The root-tip, growing of course downwards, pressed against a light wooden arm projecting from the upper plate. It was found that the maximum deflexion of the spring due to the whole extent of growth observed could be produced by a force of a few milligrammes, and thus the pressure on the plant only amounted to a few dynes. The growth pulses noticed by Bose* are shown very convincingly. Check experiments under closely similar conditions (with *dead* plants) show absolutely no recorded motion. Unfortunately, the two students of botany who kindly assisted me in this plant-growth work—Miss Cannon and Mr. Saunders—are no longer available, and this work has had to be postponed.

One of the earliest applications I made of the device was to record the minute flexure of the diaphragm of a pressure gauge. The (india-rubber) diaphragm carried one of the condenser plates; the other was mounted on a bridge across the air chamber, the usual adjustment screws being provided. My aim was to obtain a very robust pressure gauge, and I employed stout sheet rubber (1 mm. thick) stretched tightly across the air chamber (10 cm. diameter). To facilitate adjustments a small variable condenser was connected in *parallel* with the pressure-gauge condenser. I found that *both* sides of the gauge had to be enclosed in view of the high sensitivity. With one of these (*thick rubber*) I found the readings quite steady: the zero-keeping qualities and rapidity of action very high; while consistent readings could be obtained to less than 1/3 of one-millionth of an atmosphere (0·32 dyne per cm.^{2}, to be exact). In this particular case the galvanometer was still shunted 10-fold, which was necessary for two reasons: (1) as usual, no screening was employed; (2) the pressures being measured were due to the pressure drop accompanying the slow flow of air through a 1-cm. diameter (50 cm. long) brass tube, and it was impossible to avoid slight irregularities in the air stream.

For the information of those interested in gas-flow measurement, I may here mention that the slowest velocity actually dealt with was 14 cm. per sec., which produced 31 mm. deflexion on the galvanometer, while the latter was consistent to 1 mm. with the 10 shunt used. With a Pitot tube (for which pressure varies as square of air

* Bose, Phil. Trans. Roy. Soc. B. vol. xc. p. 364, Feb. 1919.

velocity) I found that an air stream of about 80 cm. per second could be measured with about the same accuracy. These numbers are *conservative statements* of results obtained *under ordinary laboratory conditions.*

I have not hitherto tried the micrometer for indicating the changes in level of liquids, but there does not appear to be any reason why it should not be so employed. In this way, using mercury, it should be possible to record 10^{-7}cm. (pressure difference) without difficulty, or, with other liquids, perhaps lower differences still. Special difficulties are presented by the problem of high vacuum pressures, but the manifest advantages of a direct-reading gauge for these encourages one to persevere in this direction.

Interesting possibilities arise if we consider the incorporation of the ultramicrometer as the recording device in a spring balance. In other brief publications, I have indicated how it can be employed to dispense with the small weights and riders of a sensitive balance. Rapidity of action and very high sensitivity are in this case no longer incompatible. However, except for special work, I imagine that most users of balances would look askance on the necessary batteries, galvanometer, and other apparatus. I have constructed an instrument, however, by which it is possible to weigh 200 grammes to the nearest 1/10 milligramme in about a quarter of a minute.

The principle employed in these balances has other interesting applications, and it will be worth while explaining it more fully. Suppose Hooke's law to be obeyed. Then $Se = Mg$, where e = total extension due to mass M and S is a constant. Small changes in e are indicated by attaching one of our condenser plates to the extremity of the spring. In a balance the load is kept within a fraction of a gramme (say) of a standard load M by removing weights when placing the body in position. The galvanometer indicates by how many milligrammes (say) the total load differs from M.

The same device should also indicate changes in g. Here there are three cases to be considered: (1) variation of terrestrial gravity from place to place; (2) variation of the vertical gravitational force due to the tide-generating forces of sun and moon; (3) "Newtonian attraction" experiments due to the introduction of a heavy mass beneath. The last-mentioned type of problem is, comparatively, a simple proposition. With a view to testing its possibility, I set up an apparatus in which a weight of about 1 kilo was suspended by a long silk cord from a

steel spring. The spring extension was about 25 cm. and the condenser plate was attached to the spring, the silk cord passing down through a hole pierced in the lower (fixed) condenser plate. The silk cord enabled the electrical circuit to be shielded from the possible capacity effects accompanying the movements of the large attracting mass of about 50 kilos of lead. The sensitivity required was very high, a simple calculation showing that, if the attracting masses were about 25 cm. apart, the extension to be expected is only 1/8,000,000 cm. By working late at night and remaining almost motionless during the observations, I was able to secure a sensitiveness such that the galvanometer showed a deflexion of about 300 scale-divisions each time the heavy mass was wheeled into place. My purpose in making this experiment was, principally, to ascertain whether the electrostatic attraction between the condenser plates would cause instability in the system. No sign of this effect was perceived.

The two other gravitational effects are much larger than that dealt with in the experiment, but involve special difficulties. These arise in both cases from the fact that the micrometer would be called on to measure the displacement occurring at long intervals of time. I have not yet been able to carry out satisfactorily experiments to test the possibility of determinations of this kind.

The ultramicrometer lends itself specially well as a recorder for a seismometer. Thanks to the kindness of the Rev. W. O'Leary, S.J., who constructed a new type of vertical seismometer for me, I was enabled to carry out some trials in this connexion. I am now, however, conducting further experiments on this matter, and shall reserve a full account of all the work for a later communication.

Perhaps the most important application of the ultramicrometer is one on which I am still engaged. A full account will, I hope, be ready for publication shortly. Thanks to the great rapidity with which the oscillatory circuit responds to the movements of the condenser plate, it is possible to use the apparatus to measure the rapidly varying stresses in machinery. Thus the out-of-balance forces on the bearings of a rotor can be determined by a suitable adaptation of the method, and the necessary data obtained for correcting the out-of-balance condition. In this connexion the use of a high-frequency oscillograph might be employed, but there are serious objections to this. Such instruments are very insensitive, and necessitate the

use of power valves operating at perhaps 1000 volts. I was enabled to carry out some experiments with ultramicrometer circuits, employing such high-power valves with a Duddell oscillograph, thanks to the courtesy of the Cambridge-Paul Company, who lent me the latter instrument. Apart from its comparative insensitiveness, the apparatus was otherwise satisfactory, but both dangerous and troublesome to separate. For these reasons, I have abandoned this method of attack for one in which an ordinary slow-period, sensitive galvanometer is employed in conjunction with a synchronous contact device. This arrangement has now been in use for some months and has fulfilled every requirement.

In this paper I have confined myself to an account of those applications of the ultramicrometer actually realized, and have avoided quoting any observations with the apparatus in such a super-sensitive condition as to render shielding and other precautions necessary. I hope, in due course, to take up this aspect of the problem, and to determine the highest working sensitivity obtainable under specially favourable conditions. One of the most serious difficulties which has to be dealt with when working at very high sensitivity is the tendency of the galvanometer to "creep." Great care is necessary in selecting the cells (accumulator and high-tension batteries) to ensure that they are in good condition. It is *absolutely necessary* to use for the zero shunt cells of identical construction to those in the H.T. anode battery. Temperature fluctuations, however small, must be avoided. Even barometric and humidity variations may conceivably affect the coils or condenser. It must be remembered how small a variation in the circuit constants will affect the anode current to a measurable extent, and seemingly unimportant factors should be carefully considered. Finally, for the higher sensitivity, one must guard against effects due to the creeping of an electrostatic charge on the interior of the valve bulb. If this cannot be prevented, at least one must wait long enough for a steady condition to be attained. The only general rule that can be given is to make the apparatus function as near the foot of the "capacity-current" curve as possible; by so doing, the minimum demand is made on the batteries, and they are less likely to change. Similarly, it is better to seek sensitivity by using a good galvanometer than by running the filament at a higher temperature. I generally use only 4 volts.

VII. *A Note on the Fluctuation of Water-Level in a Tidal-Power Reservoir.*

By Prof. S. CHAPMAN, *M.A., D.Sc., F.R.S., University of Manchester* *.

§ 1. THIS note relates to the variation of water-level in a tidal basin separated from the sea or from an estuary by a dam, in which there are sluices constantly submerged, through which water may freely flow in and out. Corresponding to a known periodic variation of level in the sea there will result a periodic rise and fall of level in the basin, and it is sought to determine the amount and character of this fluctuation. The problem is of some practical interest, since a periodic difference of height between such a basin and the sea may be used for the development of water-power, during the periods both of inward and outward flow. The insertion of turbines in the sluices modifies the rate of flow of water, but the consideration of the results for the open sluices is the first step towards ascertaining the available power.

The most important case is that in which H, the height of water in the sea, does not differ much from a simple sine-function of the time, with a period equal to half a lunar day, or 12 hours 25 minutes. Thus, let

$$H = H_0 \sin \alpha t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where H_0 is half the tidal range, t is time reckoned in seconds from the epoch of mean level during the rise of tide, and

$$\alpha = 2\pi / 12\tfrac{25}{60} \times 60 \times 60 \quad . \quad . \quad . \quad . \quad (2)$$

The height of water in the basin will be denoted by H′, while the difference of height (the working head) will be denoted by h, i.e.

$$h = H - H'. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The velocity of flow, v, through the sluices will be calculated by the usual approximate formula

$$v^2 = 2g\,|h|, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where h denotes the numerical (positive) value of h, and $g = 32{\cdot}2$.

Let a, A denote respectively the sectional area of the combined sluices and of the basin, A being assumed the same

* Communicated by the Author.

for all depths H'. Then the differential equation connecting H and H' or h is

$$\frac{d(\mathrm{AH}')}{dt}=av=\pm a(2g|h|)^{1/2} \quad . \quad . \quad . \quad . \quad (5)$$

or

$$\frac{d\mathrm{A}(\mathrm{H}-h)}{dt}=\pm a(2g|h|)^{1/2}, \quad . \quad . \quad . \quad . \quad (6)$$

where the sign of the last term is that of h.

An exact general solution of this equation, when H is an arbitrary periodic function of the time, must be difficult to obtain, on account of the transference of the sign of h, on the right-hand side, to the outside of the radical $(2g|h|)^{1/2}$. The problem was proposed to the writer some years ago, without result, by Prof. A. H. Gibson, who thereafter solved it in a number of special cases by a convenient graphical method based partly on successive trial. In these cases, which seem to lie within the range of values of most importance in practice, H was either a pure sine function, or nearly so (the latter case related to spring tides, in which the rise of level took place rather more quickly than the fall); the fluctuation within the basin proved to be a close approximation to a sine curve, the higher harmonic components in H' being very small. This result suggested to the writer the following treatment of the problem, which leads to a convenient expression for the leading component in H' or h, agreeing within two or three per cent. with the results found by Prof. Gibson, and obviating the special graphical solution of each particular case.

§ 2. It is assumed that H' or h can be represented approximately by a simple sine curve, viz.,

$$h=h_0 \sin(\alpha t+\epsilon), \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where h_0 is positive. In this case it is not difficult to represent $\pm|h|^{1/2}$, where the sign is that of h itself, by a Fourier's series. In so doing, it is convenient, for the moment, to neglect the phase-constant ϵ, and to write

$$\pm|\sin\alpha t|^{1/2}=a_1\sin\alpha t+a_3\sin 3\alpha t+a_5\sin 5\alpha t+\ldots, \quad . \quad (8)$$

since it is evident that only sine terms involving odd multiples of αt will appear in the series. The coefficient

a_n is given by

$$a_n = \frac{4\alpha}{\pi}\int_0^{\pi/2\alpha} (\sin \alpha t)^{1/2} \sin n\alpha t \, dt = \frac{4}{\pi}\int_0^{1/2\pi} (\sin x)^{1/2} \sin nx \, dx, \quad . \quad . \quad . \quad . \quad (9)$$

which may be evaluated in general terms by the expansion of $\sin nx$ in a series of products of sines and cosines of x. It is sufficient, however, to determine the first two or three terms: thus, writing $\sin x = u^{1/2}$,

$$a_1 = \frac{4}{\pi} \cdot \frac{1}{2}\int_0^1 u^{1/4}(1-u)^{-1/2} du = \frac{2}{\pi} \mathrm{B}\left(\frac{5}{4}, \frac{1}{2}\right) = \frac{2}{\pi} \frac{\Gamma(\frac{5}{4})\Gamma(\frac{1}{2})}{\Gamma(\frac{7}{4})}, \quad . \quad . \quad . \quad . \quad (10)$$

where

$$\Gamma(\tfrac{1}{2}) = \pi^{1/2}, \quad \Gamma(\tfrac{5}{4}) = 0{\cdot}9063, \quad \Gamma(\tfrac{7}{4}) = 0{\cdot}9189, \quad . \quad (11)$$

so that

$$a_1 = \frac{1}{\pi^{1/2}} \frac{0{\cdot}9063}{0{\cdot}9189} = 1{\cdot}113. \quad . \quad . \quad . \quad . \quad (12)$$

Similarly,

$$a_3 = \frac{2}{\pi}\int_0^1 (3u^{1/4} - 4u^{5/4})(1-u)^{-1/2} du = 3a_1 - 4 \, . \, \tfrac{5}{7}a_1 = \tfrac{1}{7}a_1, \quad . \quad (13)$$

$$a_5 = a_1\left(5 - 20 \, . \, \tfrac{5}{7} + 16 \, . \, \frac{9 \, . \, 5}{11 \, . \, 7}\right) = \frac{5}{77}a_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and so on. Consequently

$$\pm |\sin \alpha t|^{1/2} = 1{\cdot}113(\sin \alpha t + \tfrac{1}{7} \sin 3\alpha t + \tfrac{5}{77} \sin 5\alpha t + \ldots.), \; . \quad (15)$$

while

$$\pm(2g|h|)^{1/2} = \pm(2gh_0|\sin(\alpha t + \epsilon)|)^{1/2}$$
$$= 1{\cdot}113(2gh_0)^{1/2}\{\sin(\alpha t + \epsilon) + \tfrac{1}{7} \sin 3(\alpha t + \epsilon) + \ldots.\}, \; . \quad (16)$$

Hence, if H is of the form (1), the equation (6) may be written

$$\alpha \mathrm{A}\{\mathrm{H}_0 \cos \alpha t - h_0 \cos(\alpha t + \epsilon)\} = 1{\cdot}113(2gh_0)^{1/2}\{\sin(\alpha t + \epsilon) + \tfrac{1}{7} \sin 3(\alpha t + \epsilon) + \ldots.\}, \quad . \quad (17)$$

or, writing

$$t' = \alpha t + \epsilon, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

(6) becomes

$$\mathrm{H}_0 \cos(t' - \epsilon) - h_0 \cos t' = (2kh_0)^{1/2}(\sin t' + \tfrac{1}{7} \sin 3t' + \ldots.), \; (19)$$

where

$$(2k)^{1/2} = 1{\cdot}113 \frac{(2g)^{1/2}}{\alpha} \frac{a}{\mathrm{A}},$$

or, if α is given by (2),

$$k = 2{\cdot}018 \, . \, 10^9 (a/\mathrm{A})^2. \quad . \quad . \quad . \quad . \quad (20)$$

The equation (19) can be satisfied by suitable choice of h_0 and ϵ, provided the terms representing the third and higher components of $\pm h|^{1/2}$, on the right, are negligible. Whether this is the case, or not, is a matter for subsequent examination (§5); assuming it to be so, (19) is satisfied if

$$h_0 = H_0 \cos\epsilon, \qquad H_0 \sin\epsilon = (2kh_0)^{1/2}. \quad . \quad . \quad (21)$$

Hence, by squaring and adding,

$$H_0^2 = h_0^2 + 2kh_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

so that, remembering that h_0 is positive,

$$h_0 = (H_0^2 + k^2)^{1/2} - k. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

After thus determining h_0, the phase angle ϵ is given by

$$\tan\epsilon = (2k/h_0)^{1/2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

The equations (23, 24) determine the approximate value of the "working head" h, i.e., the difference of level on the two sides of the dam. The height H′ inside the basin is then given by

$$\left.\begin{aligned} H' = H - h &= H_0 \sin\alpha t - h_0 \sin(\alpha t + \epsilon) \\ &= (H_0 - h_0\cos\epsilon)\sin\alpha t - h_0 \sin\epsilon\cos\alpha t \\ &= H_0' \sin(\alpha t - \theta) \end{aligned}\right\}, \quad . \quad . \quad (25)$$

where

$$H_0'\cos\theta = H_0 - h_0\cos\epsilon, \qquad H_0'\sin\theta = h_0\sin\epsilon. \quad . \quad . \quad (26)$$

Thus

$$\left.\begin{aligned} H_0'^2 &= H_0^2 + h_0^2 - 2H_0h_0\cos\epsilon \\ &= H_0^2 - h_0^2 = 2kh_0 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

by (21), (22), while

$$\tan\theta = \frac{h_0\sin\epsilon}{H_0 - h_0\cos\epsilon} = \frac{h_0\sin\epsilon\cos\epsilon}{H_0\cos\epsilon - h_0\cos^2\epsilon} = \frac{\sin\epsilon\cos\epsilon}{1-\cos^2\epsilon} = \cot\epsilon,$$

so that . . . . . (28)

$$\theta = \frac{\pi}{2} - \epsilon, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

and

$$H' = (2kh_0)^{1/2}\cos(\alpha t + \epsilon). \quad . \quad . \quad . \quad . \quad . \quad (30)$$

These results also follow easily from a vector diagram; if vectors are drawn to represent H_0 and h_0, the latter being in advance of H_0 by an angle ϵ, the equation $h_0 = H_0\cos\epsilon$ indicates that the vector $H_0 - h_0$, joining the ends of the vectors H_0 and h_0, is perpendicular to h_0 (i.e., it has a lag

equal to $\frac{1}{2}\pi-\epsilon$), and that its magnitude is $H_0 \sin\epsilon$ or $(2kh_0)^{1/2}$.

§ 3. These results may be compared with those obtained graphically by Prof. Gibson for certain special cases*. The first series of these refer to the case when

$$H = 8 \sin nt. \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

where n differs slightly from α, as given in § 2 (2), because the period assumed was 13 hours: the corresponding value of k is $2{\cdot}214 \,.\, 10^9(a/A)^2$. The values of A/a considered by Prof. Gibson were 16000, 20000, 25000, and 30000. His curves for H' have been harmonically analysed, and the values of H_0' and θ for the leading component thus found; from them the semi-amplitude h_0 and phase ϵ of the leading component of h ($=H-H'$) have been calculated; they are given below, together with the values of k, h_0, and ϵ determined by the formulæ of § 2. The relations of § 2 between H_0, H_0', h_0, ϵ, and θ are not exactly fulfilled by the graphically determined values, but the agreement between the following two sets of values of h_0 (the most important quantity in practice) is satisfactory.

A/a	16000	20000	25000	30000
k	8·65	5·53	3·54	2·46
h_0 (calc.) ..	3·14	4·19	5·21	5·91
h_0 (graph.) ..	3·17	4·30	5·21	5·84
ϵ (calc.)	67°	58°	49°	42°
ϵ (graph.) ..	62	55	47	41

In the second series of curves given by Prof. Gibson, the period was taken as $12\frac{1}{2}$ hours, and H was no longer represented by a pure sine curve; the curve represented the spring tides at Portishead in the Severn Estuary, and on harmonic analysis (taking the phase of the first component as zero) the following were found to be the leading terms in its Fourier's series:

$$21{\cdot}0 \sin \alpha t + 2{\cdot}4 \sin 2\alpha t + 0{\cdot}66 \sin 3\alpha t. \quad . \quad . \quad . \quad (32)$$

Two values of A/a were considered, viz., 12,000 and 16,000. Proceeding as before, the following results were obtained by

* ‘Engineering,’ Dec. 17, 1920, p. 794.

analysis of Prof. Gibson's graphs of H′, and by calculation from the formulæ of § 2.

A/a	12000	16000
h	14·21	7·99
h_0 (calc.) ..	11·1	14·5
h_0 (graph.)..	11·4	15·2
ϵ (calc.)....	58°	46°
ϵ (graph.) ..	53	43

The differences between the two sets of values of h_0 are here slightly greater than before, though still less than 5 per cent.

Sluice area for maximum power.

§ 4. The power obtainable by using the inward and outward flow of water through the sluices is proportional (i.) to the sluice area a, and (ii.) to the mean square velocity of the water, *i. e.* to the mean value of $|h|$. When h is a pure sine-function, or nearly so, the semi-amplitude being h_0, the power available is therefore proportional to ah_0, and it is of interest to determine what ratio of sluice-area to basin-area, for a given basin-area A and tidal range $2H_0$, gives the maximum power. Since a/A occurs only in h, which is proportional to $(a/A)^2$, the condition for maximum power is clearly

$$\frac{dk^{1/2}h_0}{dk}=0,$$

or

$$\frac{dk^{1/2}\{(k^2+H_0^2)^{1/2}-k\}}{dk}=0,$$

or

$$\frac{1}{2k^{1/2}}\{(k^2+H_0^2)^{1/2}-k\}+k^{1/2}\left\{\frac{k}{(k^2+H_0^2)^{1/2}}-1\right\}=0,$$

which reduces to

$$k=\tfrac{1}{3}H_0, \quad \text{. . . . . (33)}$$

or

$$\frac{a}{A}=\frac{\alpha H_0^{1/2}}{1{\cdot}113(3g)^{1/2}}. \quad \text{. . . . . . (34)}$$

The most economical ratio of sluice-area to basin-area is thus proportional to the square root of the tidal range.

With the value of α corresponding to a lunar semi-diurnal tide (*cf.* (2)) the above result is equivalent to

$$\frac{A}{a}=7{\cdot}79\,.\,10^4H_0^{-1/2},$$

so that if $H_0 = 21$ ft., $A/a = 17{,}000$. In the case of the first set of Prof. Gibson's curves, where the period of the tide was taken as 13 hours, the corresponding value of A/a would be $8{\cdot}15 \,.\, 10^4 H_0^{-1/2}$, so that if $H_0 = 8$, the best value of A/a is 28,800. By inference from his graphical results, Prof. Gibson estimated these two ratios of A/a as 16,000 and 30,000 respectively.

§ 5. It remains to consider how far the neglect of the higher harmonic terms in h is likely to affect the estimate of the available power. The coefficients of the neglected third and fifth harmonics in $h^{1/2}$ in (19) are 1/7 and 5/77 times the coefficient of the leading term in $|h|^{1/2}$, and it would seem as if the odd harmonics in the true value of h bear smaller ratios than these to the leading term, while the even harmonics are small, if not zero. The coefficients of the second harmonics in H' or h' corresponding to the four values of A/a in Prof. Gibson's first set of curves are 0·03, 0·04, 0·00, 0·06, while the corresponding coefficients of the third harmonics are 0·39, 0·28, 0·57, 0·47; part of these may be due to slight errors in measuring the curves. Now the utmost extent to which the mean value of $|h|$ can be affected by a harmonic of frequency n and amplitude a_n is easily seen to be a_n/n. In the cases cited this is quite negligible. Even for the spring-tide curves discussed in § 3, where H was itself not a pure sine-function (*cf.* 32) the second and third harmonic terms in h, as derived graphically, were only 1·47 and 0·70 ($A/a = 12{,}000$) and 0·70, 0·63 ($A/a = 16{,}000$). The second harmonics would in these cases affect the mean value of h very little, because the periods of inflow and outflow for the reservoir are very nearly equal. Thus, though the analysis of this paper could be extended, if necessary, so as to afford an approximation to at least the next most important harmonic component in h after the first, there appears to be no practical necessity for so doing.

Summary.

§ 6. The difference of height, h, between (i.) a sea or estuary undergoing a tidal change of water-level, in half a lunar day, represented by $H_0 \sin \alpha t$, and (ii.) a tidal basin of area A connected with the sea by sluices of area a, is represented by the approximate formula

$$h = h_0 \sin(\alpha t + \epsilon),$$

where h_0 and ϵ are given by

$$h=(H_0^2+h^2)^{1/2}-k, \quad \tan\epsilon=(2k/h_0)^{1/2}$$

in terms of the number k, where

$$k=2{\cdot}018\,.\,10^9\,(a/A)^2.$$

The ratio of A to a which corresponds to the maximum power-development from this difference of head is given by

$$A/a=7{\cdot}79\,.\,10^4\,H_0^{-1/2},$$

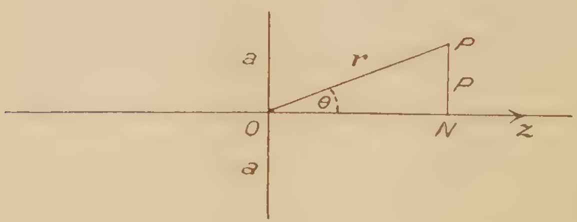
where H_0 is supposed measured in feet.

VIII. *Note on the Magnetic Field produced by Circular Currents.* *By* T. J. I'A. BROMWICH, *Sc.D., F.R.S., Fellow and Prælector in Mathematical Science, St. John's College, Cambridge* *.

PROF. H. NAGAOKA has investigated these magnetic fields by means of q-series in a recent number of this journal † ; but it does not seem to have been noticed that his results can be found in a very elementary way by using Legendre's theorem that the potential of a symmetrical system can be determined when the form of the potential is known along the axis of symmetry.

1. *The field of a single circle (of radius a).*—At points on

Fig. 1.



the axis of z the magnetic force due to unit current in the circular wire is equal to

$$\gamma=\frac{2\pi a^2}{(a^2+z^2)^{3/2}}=\frac{2\pi}{a}\left(1-\frac{3}{2}\frac{z^2}{a^2}+\frac{3\,.\,5}{2\,.\,4}\frac{z^4}{a^4}-\dots\right),$$

provided that z is numerically less than a.

* Communicated by the Author.
† Phil. Mag. vol. xli. p. 377 (1921).

Now the component of magnetic force in any fixed direction satisfies Laplace's equation: and in particular this is true of γ, the component parallel to the axis of symmetry.

Thus at points not on the axis the magnetic force γ is given by the series

$$\gamma=\frac{2\pi}{a}\left\{1-\frac{3}{2}\frac{r^2}{a^2}P_2(\cos\theta)+\frac{15}{8}\frac{r^4}{a^4}P_4(\cos\theta)-\ldots\right\}$$

$$=\frac{2\pi}{a}\left\{1-\frac{3}{2a^2}\left(z^2-\frac{1}{2}\rho^2\right)+\frac{15}{8a^4}\left(z^4-3z^2\rho^2+\frac{3}{8}\rho^4\right)-\ldots\right\}. \quad (1)$$

The formula (1) agrees with Prof. Nagaoka's (18), (19): to obtain (20) we must include the following term in the bracket, which is

$$-\frac{35r^6}{16a^6}P_6(\cos\theta)=-\frac{35}{16a^6}\left(z^6-\frac{15}{2}z^4\rho^2+\frac{45}{8}z^2\rho^4-\frac{5}{16}\rho^6\right). \quad (2)$$

It will be observed that the coefficient of ρ^6 is apparently erroneous in Prof. Nagaoka's formula.

The value of the magnetic potential V can now be found by integrating the equation

$$-\frac{\partial V}{\partial z}=\gamma.$$

This gives, at points on the axis,

$$V_0-V=2\pi\left(\frac{z}{a}-\frac{1}{2}\frac{z^3}{a^3}+\frac{1.3}{2.4}\frac{z^5}{a^5}-\frac{1.3.5}{2.4.6}\frac{z^7}{a^7}+\ldots\right),$$

where V_0 is a constant: and so, at any point, we have the formula

$$V_0-V=2\pi\left(\frac{r}{a}P_1-\frac{1}{2}\frac{r^3}{a^3}P_3+\frac{3}{8}\frac{r^5}{a^5}P_5-\frac{5}{16}\frac{r^7}{a^7}P_7+\ldots\right).$$

When the last formula is written out in full we obtain the result

$$V_0-V=2\pi\left\{\frac{z}{a}-\frac{1}{2a^3}\left(z^3-\frac{3}{2}z\rho^2\right)+\frac{3}{8a^5}\left(z^5-5z^3\rho^2+\frac{15}{8}z\rho^4\right)\right.$$
$$\left.-\frac{5}{16a^7}\left(z^7-\frac{21}{2}z^5\rho^2+\frac{105}{8}z^3\rho^4-\frac{35}{16}z\rho^6\right)\right\}. \quad (3)$$

Thus on differentiation with respect to ρ we find λ the transverse component of the magnetic field

$$\lambda=\sqrt{(\alpha^2+\beta^2)}=-\frac{\partial V}{\partial\rho}=\frac{2\pi z\rho}{a^3}\left\{\frac{3}{2}+\frac{15}{16a^2}(3\rho^2-4z^2)\right.$$
$$\left.+\frac{105}{128a^4}(8z^4-20z^2\rho^2+5\rho^4)+\ldots\right\}. \quad (4)$$

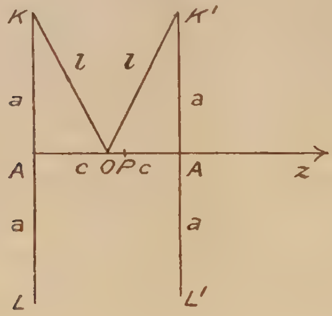
The formula (4) agrees precisely with (15), (16), and (17) of Prof. Nagaoka's.

2. *Field near the centre of Helmholtz's galvanometer (double coils).*

Let the section be represented by the adjoining diagram (fig. 2) in which

$$a=2c, \text{ and } l^2=a^2+c^2=5c^2.$$

Fig. 2.



$2c=a, \quad l^2=a^2+c^2=5c^2$

Since the field is symmetrical about the line of centres AA′ (which we take as Oz), it will suffice to calculate the field at a point P on the axis (and near to O).

The magnetic force at the point P due to unit current in the coil A is known to be

$$\gamma_1=\frac{2\pi a^2}{\mathrm{KP}^3}=\frac{2\pi a^2}{\{a^2+(c+z)^2\}^{3/2}}.$$

Now we have

$$a^2+(c+z)^2=l^2+2cz+z^2=\left(l+\frac{cz}{l}\right)^2+\left(\frac{az}{l}\right)^2.$$

Thus, since az/l^2 is small, we can write

$$\gamma_1=\frac{2\pi a^2}{\{a^2+(c+z)^2\}^{3/2}}=\frac{2\pi a^2 l^3}{(l^2+cz)^3}\left\{1+\left(\frac{az}{l^2+cz}\right)^2\right\}^{-3/2}$$

$$=\frac{2\pi a^2}{(l^2+cz)^3}\left\{1-\frac{3}{2}\frac{a^2z^2}{(l^2+cz)^2}+\frac{15}{8}\frac{a^4z^4}{(l^2+cz)^4}+\dots\right\}.$$

To obtain the same degree of accuracy as would be given

by Prof. Nagaoka's formulæ, we must expand as far as terms in z^4. Then the result is

$$\gamma_1 = \frac{2\pi a^2}{l^3}\left(1 - \frac{3cz}{l^2} + \frac{6c^2z^2}{l^4} - \frac{10c^3z^3}{l^6} + \frac{15c^4z^4}{l^8}\right)$$
$$- \frac{3\pi a^4 z^2}{l^7}\left(1 - \frac{5cz}{l^2} + \frac{15c^2z^2}{l^4}\right) + \frac{15\pi a^6 z^4}{4l^{11}} \quad . \quad . \quad (5)$$

The magnetic force γ_2, due to the same current in the second coil, is given by changing the sign of z in γ_1; and so the complete field at P is

$$\gamma = \gamma_1 + \gamma_2 = \frac{4\pi a^2}{l^3} + \frac{6\pi a^2 z^2}{l^7}(4c^2 - a^2) + \frac{15\pi a^2 z^4}{2l^{11}}(8c^4 - 12a^2c^2 + a^4),$$

correct as far as terms in z^4.

So far we have not used the special feature of the Helmholtz arrangement: namely, the relation $a = 2c$, which provides for the disappearance of the second term in γ.

Inserting this value, we find that on the axis the magnetic field is given by the approximate formula

$$\gamma = \frac{4\pi a^2}{l^3}\left(1 - \frac{9}{5}\frac{z^4}{l^4}\right). \quad . \quad . \quad . \quad . \quad . \quad (6)$$

To obtain the field at other points we must calculate first the magnetic potential at P; this is given by

$$-\frac{\partial V}{\partial z} = \gamma.$$

and so, on integrating (6), we find that

$$V = V_0 - \frac{4\pi a^2}{l^2}\left(\frac{z}{l} - \frac{9}{25}\frac{z^5}{l^5}\right) = V_0 - \frac{16\pi}{5}\left(\frac{z}{l} - \frac{9}{25}\frac{z^5}{l^5}\right), \quad (7)$$

where V_0 is the value of V at O.

Accordingly the value of V at any other point near O is given by the approximate formula

$$V = V_0 - \frac{16\pi}{5}\left\{\frac{r}{l}P_1(\cos\theta) - \frac{9}{25}\frac{r^5}{l^5}P_5(\cos\theta)\right\}.$$

As before, we have

$$rP_1(\cos\theta) = z,$$

and
$$r^5 P_5(\cos\theta) = z^5 - 5z^3\rho^2 + \frac{15}{8}z\rho^4,$$

where
$$\rho = r\sin\theta.$$

Thus the potential is given by the approximation

$$V=V_0-\frac{16\pi}{5l}z+\frac{18\pi}{125l^5}(8z^5-40z^3\rho^2+15z\rho^4). \quad . \quad . \quad (8)$$

Hence the complete field is given by the two components

$$\lambda=\sqrt{(\alpha^2+\beta^2)}=-\frac{\partial V}{\partial \rho}=\frac{72\pi}{25l^5}z\rho(4z^2-3\rho^2) \quad . \quad . \quad . \quad . \quad (9)$$

$$\gamma=-\frac{\partial V}{\partial z}=\frac{16\pi}{5l}\left\{1-\frac{9}{5l^4}\left(z^4-3z^2\rho^2+\frac{3}{8}\rho^4\right)\right\} \quad . \quad . \quad . \quad (10)$$

On remembering that $l^2=\frac{5}{4}a^2$, it will be seen that formulæ (9) and (10) are the same as (C) and (D) of Prof. Nagaoka's paper (*l. c.* p. 387).

IX. *Repulsion Effect between the Poles of an Electric Arc.*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

I HAVE read with considerable interest the various theories * which have been advanced to account for the repulsion effect between the poles of an electric arc. Perhaps you will be good enough to grant me the privilege of suggesting an explanation on somewhat different lines.

Whilst the projection of particles across the gap and the so-called electric wind accompanied by its electrostatic reaction may have an influence on the forces between the electrodes, it seems to me that the "pinch effect" in the arc must also be an important factor. The work of Bary † and Northrup ‡ on electromagnetic striction in mobile conductors has left little room for doubt regarding the generality of this phenomenon. The characteristic interruptions and stratification in vacuum-tube discharges, the intermittent current taken

* A. Sellerio, *N. Cimento*, xi. p. 67 (1916) & Phil. Mag. (6) xliv. p. 765 (1922). W. G. Duffield, T. H. Burnham, and A. H. Davies, Roy. Soc. Phil. Trans. 220. p. 109 (1919). W. G. Duffield and Mary D. Waller, Phil. Mag. xl. p. 781 (1920). S. Ratner, Phil. Mag. xl. p. 511 (1920). A. M. Tyndall, Phil. Mag. xl. p. 780 (1920), & xlii. p. 972 (1921). H. E. G. Beer and A. M. Tyndall, Phil. Mag. xlii. p. 956 (1921).

† P. Bary, *L'clairage Électrique*, li. p. 37 (1907); *Lumière Électr.* vi. p. 135 (1909); *Le Radium*, iv. p. 323 (1907); *Journ. de Physique*, viii. p. 190 (1909).

‡ E. F. Northrup, Phys. Rev. xxiv. p. 474 (1907).

by mercury-vapour lamps working on a D. C. supply, the singing and whistling arcs, can all be simply explained by this action.

A conductor may be imagined to consist of a large number of filaments parallel to the axis, each carrying part of the total current. All the filaments except the one on the axis will experience a force tending to urge them towards the centre. If, therefore, the conducting column is mobile, it will contract so as to diminish its cross-section and, in consequence, increase in length. If the column cannot increase in length, the hydrostatic pressure will become greater along the axis than near the surface. This conception is easily developed mathematically for a cylindrical conductor. In such a case the intensity of the pressure (p) at any radius (r) is given by

$$p = \frac{I^2}{\pi R^4}(R^2 - r^2),$$

where I = total current and R = external radius of the conductor.

Bary has clearly demonstrated an increase of pressure, towards the axis of a conducting column of gas.

Although a gas will not normally resist change of pressure we have to remember that in the case of an arc the radial striction forces are exerted more or less uniformly along its whole length, so that the ionized vapour is kept within a confined space. The system is analogous to a rubber tube containing gas under pressure and having rigid plugs in its ends. In liquids the contraction is generally more pronounced about midway between the electrodes on account of the relatively large surface friction. In gases the effect will be chiefly a volume change and more uniformly distributed.

If we assume that the conducting column of the arc is cylindrical and contracts uniformly, the force (F) on each pole due to the "pinch effect" is obtained by integrating the pressure over the section. For an equilibrium condition it is easy to show that

$$F = \frac{I^2}{2},$$

where I = total current.

The experimental results obtained by Prof. Duffield suggest that the total force on each electrode, after making proper allowance for the influence of the earth's field and the electrodynamic action of the circuit, is approximately proportional to the square of the current. Moreover, taking a

current of 20 amperes, theoretically we should get about 2 dynes on each pole from this action, which amounts to nearly 30 per cent. of the observed force. It is also significant that the effect is the same for alternating and direct currents.

The researches of Beer and Tyndall on an arc between carbons drilled centrally showed that the hydrostatic pressure is greatest along the axis of the conducting vein.

The observed increased force on the electrodes when an arc is hissing or unsteady might also be expected on the hypothesis that electromagnetic striction is largely responsible. At the moment when the conducting column is ruptured the pressure is very high, and there is a quick projection of the particles to the right and left of the contracted section, producing an increased force on the poles. In some cases the displacement velocity must be very large and the resulting instantaneous pressure correspondingly high. Trotter * has shown that parts of an arc are in rapid motion during the unstable state when hissing begins.

I suggest that the evidence is in favour of the view that at least a measurable part of the repulsion effect in an arc is due to the "pinch" phenomenon operating in the gaseous conducting column between the poles.

I am, Gentlemen,

Yours faithfully,

H. Monteagle Barlow.

X. *The Effect of a Steady Wind on the Sea-level near a Straight Shore.* *By* Harold Jeffreys, *M.A., D.Sc., Fellow of St. John's College, Cambridge. Being part of a Dissertation commended by the Adjudicators for the Adams Prize*, 1923 †.

IT is a well-known result in meteorology that in any steady motion of the atmosphere, apart from the influence of friction, the wind is along the isobars and perpendicular to the horizontal thrust upon it, the higher pressure being on the right in the northern hemisphere, on the left in the southern. The same proposition, subject to the same conditions, must hold in the ocean. It has been shown ‡ that a closely related result holds in the case of ocean currents driven by the

* Trotter, Proc. Roy. Soc. lvi. p. 262 (1894).

† Communicated by the Author.

‡ Walfrid Ekman, *Arkiv. f. Matematik, Acad. Stockholm*, ii. (1905). Harold Jeffreys, Phil. Mag. xxxix. pp. 578–586 (1920).

friction of the wind blowing over the water surface; the resultant drift of the water, if unobstructed in any way, is perpendicular to the wind. The surface current, however, is inclined to the wind towards the high pressure side, and, provided the eddy viscosity is uniform down to a specifiable depth, the inclination is 45°.

It was pointed out to me by Mr. F. J. W. Whipple that this result needs to be modified when the motion is obstructed by a long straight coast, which prevents any resultant drift from taking place across a vertical section parallel to the coast. The modifications required are discussed in this paper. It is found that the ultimate effect of a steady wind is to produce an inclination of the surface of the water, the contour lines being parallel to the coast. The resultant drift in any vertical column is parallel to the coast, but the amounts of the slope and the drift depend on the wind velocity, the depth, and the eddy viscosity. In such a steady motion obstructed by a shore, the current at the surface is not in general at 45° to the wind.

Let the axis of x be taken parallel to the shore, that of y at right angles to it, and that of z vertically downwards. The origin is midway between the undisturbed free surface and the bottom, the latter being assumed horizontal. Let n be the component of the earth's angular velocity about the downward vertical. The sense of the axes is to be such that n is positive. Thus if the axis of x is eastwards, that of y is towards the nearest pole; and if the axis of x is towards the equator, that of y is to the east. With sufficient accuracy n can usually be treated as a constant.

Let p be the pressure, ρ the density, k the coefficient of eddy viscosity, and u and v the components of horizontal velocity. The equations of steady motion of the water are

$$\frac{1}{\rho}\frac{\partial p}{\partial x}=k\frac{\partial^2 u}{\partial z^2}+2nv, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\frac{1}{\rho}\frac{\partial p}{\partial y}=k\frac{\partial^2 v}{\partial z^2}-2nu. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Taking a new variable w equal to $u+iv$, we can combine these into the single equation:

$$k\frac{\partial^2 w}{\partial z^2}-2niw=\frac{1}{\rho}\left(\frac{\partial p}{\partial x}+i\frac{\partial p}{\partial y}\right). \quad . \quad . \quad . \quad (3)$$

If the elevation of the free surface above the undisturbed level is ζ, we have:

$$\frac{\partial p}{\partial x}=g\rho\frac{\partial \zeta}{\partial x}; \quad \frac{\partial p}{\partial y}=g\rho\frac{\partial \zeta}{\partial y}. \quad . \quad . \quad . \quad . \quad . \qquad (4)$$

We require to know whether a solution is possible, such that all the quantities involved are independent of x, so that $\partial p/\partial x$ is zero. With this condition (2) becomes

$$k\frac{\partial^2 w}{\partial z^2}-2niw=2niG, \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

where

$$G=-\frac{g}{2n}\frac{\partial \zeta}{\partial y}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (6)$$

On the analogy with the geostrophic wind we may think of G as the geostrophic current. G is independent of z, and therefore the solution of (5) is

$$w=G+A\sinh jqz+B\cosh jqz, \quad . \quad . \quad . \qquad (7)$$

where

$$q^2=n/k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (8)$$

$$j=1+i, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

and A and B are constants. It will be noticed that q and G are purely real; but A and B may be complex.

Let h be the depth. There must be no resultant flow across a plane parallel to the shore: for, if there were, water would be accumulating in or disappearing from the neighbourhood of the shore, and the system would not be in a steady state. Hence

$$\int_{-\frac{1}{2}h-\zeta}^{\frac{1}{2}h} v\,dz=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

Neglecting squares of small quantities, we see that $\int_{-\frac{1}{2}h}^{\frac{1}{2}h} w\,dz$ is purely real: thus $\frac{B}{jq}\sinh\frac{1}{2}jqh$ is purely real. If we call this C/q, and put $\frac{1}{2}qh=\lambda$, then (7) takes the form

$$w=G+A\sinh jqz+jC\frac{\cosh jqz}{\sinh j\lambda}. \quad . \quad . \quad . \quad . \qquad (11)$$

Actually the condition in contact with the shore must require the absence of a normal component of velocity at every point: but evidently (10) is all that can be asserted of places remote from the shore. Let the current at the bottom be flowing with speed R in a direction making the angle γ with the shore line, so that, when $z=\frac{1}{2}h$,

$$w=Re^{i\gamma}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (12)$$

Then

$$Re^{i\gamma}=A\sinh j\lambda+jC\coth j\lambda+G. \quad . \quad . \quad . \quad . \qquad (13)$$

The frictional condition for the bottom is

$$-k\rho\frac{\partial w}{\partial z}=0{\cdot}002\rho R^2 e^{\iota\gamma}, \quad . \quad . \quad . \quad . \quad (14)$$

so that, writing κ for the numerical coefficient 0·002, we have

$$-\frac{\kappa R^2}{k}e^{\iota\gamma}=jqA\cosh j\lambda+j^2qC. \quad . \quad . \quad . \quad (15)$$

Again, the condition for the free surface is

$$k\rho\frac{\partial w}{\partial z}=-\kappa\sigma Q^2 e^{\iota\alpha}, \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where $Q\cos\alpha$ and $Q\sin\alpha$ are the component velocities of the wind relative to the water at the surface, and σ is the density of the air. In general, the velocity of the wind is large compared with that of the water, so that $Q\cos\alpha$ and $Q\sin\alpha$ may be taken to be the actual components of wind velocity relative to the earth's surface.

Substituting in (16) from (11) we have

$$-\frac{\kappa\sigma}{k\rho}Q^2 e^{\iota\alpha}=jqA\cosh j\lambda-j^2qC. \quad . \quad . \quad . \quad (17)$$

Hence from (15) and (17)

$$jqA\cosh j\lambda=-\frac{1}{2}\frac{\kappa}{k}\left(R^2e^{\iota\gamma}+\frac{\sigma}{\rho}Q^2e^{\iota\alpha}\right), \quad . \quad . \quad (18)$$

$$2iqC=-\frac{1}{2}\frac{\kappa}{k}\left(R^2e^{\iota\gamma}-\frac{\sigma}{\rho}Q^2e^{\iota\alpha}\right). \quad . \quad . \quad (19)$$

And from (13)

$$\begin{aligned}G&=Re^{\iota\gamma}+\frac{1}{2}\frac{\kappa}{k}\left(R^2e^{\iota\gamma}+\frac{\sigma}{\rho}Q^2e^{\iota\alpha}\right)\frac{\tanh j\lambda}{jq}\\&\quad+\frac{1}{2jq}\frac{\kappa}{k}\left(R^2e^{\iota\gamma}-\frac{\sigma}{\rho}Q^2e^{\iota\alpha}\right)\coth j\lambda\\&=Re^{\iota\gamma}+\frac{\kappa}{qk\sqrt{2}}R^2\coth 2j\lambda e^{\iota(\gamma-\frac{1}{4}\pi)}\\&\quad-\frac{\kappa Q^2}{qk\sqrt{2}}\frac{\sigma}{\rho}\operatorname{cosech} 2j\lambda e^{\iota(\alpha-\frac{1}{4}\pi)}. \quad . \quad . \quad (20)\end{aligned}$$

If W be the velocity at the surface, we have from (11) and (13)

$$\begin{aligned}W&=Re^{\iota\gamma}-2A\sinh j\lambda\\&=Re^{\iota\gamma}+\frac{\kappa}{jkq}\tanh j\lambda\left(R^2e^{\iota\gamma}+\frac{\sigma}{\rho}Q^2e^{\iota\alpha}\right). \quad . \quad (21)\end{aligned}$$

As C is purely real, equation (19) gives

$$R^2 \cos\gamma - \frac{\sigma}{\rho} Q^2 \cos\alpha = 0. \quad . \quad . \quad . \quad . \quad (22)$$

This equation, with those obtained by equating real and imaginary parts in (20), should determine G, R, and γ in terms of Q and α. We notice that $\cos\gamma$ has the same sign as $\cos\alpha$; in other words, the components along the shore of the wind and the bottom drift are in the same direction. Equation (20) can be written

$$G = Re^{\iota\gamma} + De^{\iota(\gamma - \frac{1}{4}\pi)} \frac{\sinh 4\lambda - i \sin 4\lambda}{\cosh 4\lambda - \cos 4\lambda}$$

$$- Ee^{\iota(\alpha - \frac{1}{4}\pi)} \frac{2(\sinh 2\lambda \cos 2\lambda - i \cosh 2\lambda \sin 2\lambda)}{\cosh 4\lambda - \cos 4\lambda}, \quad (23)$$

where
$$D = \frac{\kappa R^2}{qk\sqrt{2}} = \frac{\kappa}{qk\sqrt{2}} \frac{\sigma \cos\alpha}{\rho \cos\gamma} Q^2 \quad . \quad . \quad . \quad (24)$$

$$E = \frac{\kappa}{qk\sqrt{2}} \frac{\sigma}{\rho} Q^2 = \frac{\cos\gamma}{\cos\alpha} D. \quad . \quad . \quad . \quad . \quad (25)$$

Hence

$$(G - R\cos\gamma)(\cosh 4\lambda - \cos 4\lambda)$$
$$= D\{\cos(\gamma - \tfrac{1}{4}\pi) \sinh 4\lambda + \sin(\gamma - \tfrac{1}{4}\pi) \sin 4\lambda\},$$
$$-2E\{\cos(\alpha - \tfrac{1}{4}\pi) \sinh 2\lambda \cos 2\lambda$$
$$+ \sin(\alpha - \tfrac{1}{4}\pi) \cosh 2\lambda \sin 2\lambda\}, \quad . \quad . \quad (26)$$

and

$$-R\sin\gamma(\cosh 4\lambda - \cos 4\lambda)$$
$$= D\{\sin(\gamma - \tfrac{1}{4}\pi) \sinh 4\lambda - \cos(\gamma - \tfrac{1}{4}\pi) \sin 4\lambda\}$$
$$-2E\{\sin(\alpha - \tfrac{1}{4}\pi) \sinh 2\lambda \cos 2\lambda$$
$$- \cos(\alpha - \tfrac{1}{4}\pi) \cosh 2\lambda \sin 2\lambda\}. \quad . \quad . \quad (27)$$

Substituting in (27) for D, E, and R from (22), (24), and (25), we find

$$(\cosh 4\lambda - \cos 4\lambda) \sin\gamma \left(\frac{\rho}{\sigma} \cos\alpha \cos\gamma\right)^{\frac{1}{2}} \frac{qk\sqrt{2}}{\kappa}$$

$$+ Q\left[\begin{array}{l} \cos\alpha\{\sin(\gamma - \tfrac{1}{4}\pi) \sinh 4\lambda - \cos(\gamma - \tfrac{1}{4}\pi) \sin 4\lambda\} \\ 2 - \cos\gamma \sin(\alpha - \tfrac{1}{4}\pi) \sinh 2\lambda \cos 2\lambda \\ \qquad - \cos(\alpha - \tfrac{1}{4}\pi) \cosh 2\lambda \sin 2\lambda\} \end{array}\right]$$

$$. \quad . \quad . \quad (28)$$

The square root in this equation arises as $R\cos\gamma/Q$, and evidently must have the same sign as $\cos\gamma$ and $\cos\alpha$. The equation should enable us to find γ in terms of Q and α; then (22) should give R, and finally (26) gives G. It follows that a steady motion with the contour lines parallel to the shore is in general possible. To make further progress it is necessary to consider special cases, on account of the complexity of the formulæ.

(1) *Case of deep water.*

Suppose that h is so large that $e^{-2\lambda}$, which is equal to e^{-qh}, can be neglected. Then equations (26) and (28) become

$$\sin\gamma\left(\frac{\rho}{\sigma}\cos\alpha\cos\gamma\right)^{\frac{1}{2}}\frac{\sqrt{2nk}}{\kappa}+Q\cos\alpha\sin(\gamma-\tfrac{1}{4}\pi)=0\,; \quad (29)$$

$$G=R\cos\gamma+\frac{\kappa}{\sqrt{2nk}}\frac{\sigma\cos\alpha}{\rho\cos\gamma}Q^2. \quad . \quad . \quad . \quad (30)$$

In these equations the value of q from (8) has been substituted. Substituting in (30) for R and Q in terms of γ from (22) and (29), we find

$$G=\frac{2\sqrt{nk}}{\kappa}\frac{\sin\gamma}{1-\sin 2\gamma}. \quad . \quad . \quad . \quad . \quad (31)$$

Equation (29) can be written :

$$\frac{\sin^2\gamma\cos\gamma}{1-\sin 2\gamma}=\frac{\sigma}{\rho}\frac{\kappa^2}{4nk}Q^2\cos\alpha. \quad . \quad . \quad . \quad (32)$$

Hence $$G=\frac{\kappa}{2\sqrt{nk}}\frac{\sigma}{\rho}\frac{Q^2\cos\alpha}{\sin\gamma\cos\gamma}. \quad . \quad . \quad . \quad . \quad (33)$$

We see from (29) that $\sin\gamma$ and $\sin(\gamma-\frac{1}{4}\pi)$ must have opposite signs. Hence γ lies between 0 and $\frac{1}{4}\pi$ or between π and $\frac{5}{4}\pi$. Remembering that $\cos\alpha$ and $\cos\gamma$ have the same sign, we see, therefore, that if $\cos\alpha$ is positive, γ lies between 0 and $\frac{1}{4}\pi$, and if $\cos\alpha$ is negative, γ lies between π and $\frac{5}{4}\pi$. Accordingly in the northern hemisphere, with a coast running east and west, if the wind has a component from the east, the bottom drift is towards some direction between W. and S.W.; if the wind has a component from the west, the bottom drift is towards a direction between E. and N.E. These results hold whether the other component of the wind is from the north or the south.

When γ increases from 0 to $\frac{1}{4}\pi$, $\dfrac{\sin^2\gamma\cos\gamma}{1-\sin 2\gamma}$ increases steadily from 0 to $+\infty$. Hence for every positive value of the quantity on the right of (32) there is one possible value of γ. The same is easily seen to hold for negative values. We also see that small values of $Q^2\cos\alpha$ correspond to small values of γ, while large values of $Q^2\cos\alpha$ correspond to values of γ near $\frac{1}{4}\pi$. Hence by (22) and (31) R and G tend to zero with $Q^2\cos\alpha$. A wind at right angles to the shore therefore produces no drift of the water in deep water. This agrees with previous results. In the open ocean the resultant momentum relative to the earth's surface should be at right angles to the wind, and therefore a shore at right angles to the wind should not interfere with it. A gentle wind tends to produce a general drift of the water parallel to the shore: a stronger wind at the same inclination to the shore causes the bottom drift to be markedly inclined to the shore.

Deep water: Numerical Examples.

In an ordinary case we may have $\rho=800\sigma$, $n=8\times10^{-5}$/sec., $k=100$ cm.^{2}/sec. If the wind attains the gale velocity of 18 m./s., with $\cos\alpha=1$, the quantity on the right of (32) is nearly $\frac{1}{2}$. Accordingly, for strong gales and hurricanes nearly parallel to the coast, we can treat it as large, but for lighter winds and gales nearly at right angles to the coast it is small. In the former case $\gamma=\frac{1}{4}\pi$ nearly, and in the latter γ is small.

(*a*) Hurricanes alongshore. Equation (33) reduces nearly to

$$G=\frac{\kappa}{2\sqrt{(nk)}}\frac{\sigma}{\rho}Q^2\cos\alpha, \quad . \quad . \quad . \quad . \quad (34)$$

and if Q is 40 m./s. we find that G is 4 m./s. Then

$$\frac{\partial\zeta}{\partial y}=-\frac{2nG}{g}=-\frac{6\text{ cm.}}{1\text{ km.}}.$$

Thus if such a wind was from the north, the water level along an eastern shore would be higher by 24 m. than that at an opposite shore 400 km. away.

The approximations of this paragraph are also the conditions that the square terms in (21) shall be great compared with $Re^{\iota\gamma}$; also $\tanh j\lambda$ is approximately unity.

Hence, to this accuracy,

$$W = \frac{\kappa}{\sqrt{(2nk)}}\frac{\sigma}{\rho}Q^2e^{-\frac{1}{4}\pi}\left(e^{i\gamma}\frac{\cos\alpha}{\cos\gamma}+e^{i\alpha}\right)$$

$$= \frac{\kappa}{\sqrt{(2nk)}}\frac{\sigma}{\rho}\frac{Q^2e^{-\frac{1}{4}\pi}}{\cos\gamma}\{2\cos\alpha\cos\gamma+i\sin(\alpha+\gamma)\}. \quad (35)$$

Substituting for γ we get

$$W = \frac{\kappa}{\sqrt{(2nk)}}\frac{\sigma}{\rho}Q^2e^{-\iota\frac{\pi}{4}}\{2\cos\alpha+i\cos\alpha+i\sin\alpha\}. \quad (36)$$

If the inclination of the surface-drift to the coast is β, we have

$$\beta = -\tfrac{1}{4}\pi + \tan^{-1}\tfrac{1}{2}(1+\tan\alpha)$$

$$\tan\beta = \frac{\tan\alpha - 1}{\tan\alpha+3}. \quad (37)$$

If $\alpha=0$, corresponding to a wind parallel to the shore, $\tan\beta=-\frac{1}{3}$, so that $\beta=-18°$. The surface-drift is therefore inclined towards the shore to the right of the wind, but at a smaller inclination than in the absence of any obstruction.

If $\alpha=\frac{1}{2}\pi$, $\beta=\frac{1}{4}\pi$, so that the surface-drift is at 45° to the wind. This is what we should expect, since this is the case where the coast does not interfere with the general drift.

(*b*) Light to strong wind alongshore, or gale nearly at right angles to the shore.

Here the quantity on the right of (32) is small, so that γ is small. We have then

$$\sin\gamma = Q\cos^{\frac{1}{2}}\alpha\left(\frac{\sigma}{\rho}\right)^{\frac{1}{2}}\frac{\kappa}{2(nk)^{\frac{1}{2}}}. \quad (38)$$

$$G = \left(\frac{\sigma}{\rho}\right)^{\frac{1}{2}}Q\cos^{\frac{1}{2}}\alpha. \quad (39)$$

Thus if $Q\cos^{\frac{1}{2}}\alpha = 5$ m./s., $G = 18$ cm./sec.,

and
$$\frac{\partial\zeta}{\partial y} = -\frac{0{\cdot}3\text{ cm.}}{1\text{ km.}}$$

Such a wind would therefore produce a difference of level of 1·2 metres between points at the shore and 400 km. from it.

The first term in (21) is large in this case compared with the square terms, so that

$$W = Re^{i\gamma} = G, \quad (40)$$

and as $\gamma=0$, approximately, we see that the surface-drift caused by a light wind is parallel to the shore.

(2) *Case of shallow water.*

Suppose that h is so small that λ^2, which is equal to $\frac{1}{4}q^2h^2$, can be neglected. The most convenient way of dealing with this case is to return to equation (20).

$$G-Re^{\iota\gamma}=De^{\iota(\gamma-\frac{1}{4}\pi)}\coth 2j\lambda-Ee^{\iota(\alpha-\frac{1}{4}\pi)}\operatorname{cosech} 2j\lambda. \quad (41)$$

Remembering that $e^{-\frac{1}{4}\iota\pi}=\sqrt{2}/j$, and $j^2=2i$, $q^2=n/k$,

$$E=\frac{\kappa Q^2}{qk\sqrt{2}}, \quad D=E\frac{\cos\alpha}{\cos\gamma},$$

we readily transform this to

$$G-Re^{\iota\gamma}=\frac{\kappa Q^2}{2nk}\frac{\sigma}{\rho}\frac{\sin(\alpha-\gamma)}{\cos\gamma}. \quad . \quad . \quad . \quad (42)$$

Equating imaginary parts in this equation, we see that

$$R\sin\gamma=0,$$

and therefore γ is zero or π, and the bottom drift is parallel to the shore. Equations (22) and (42) therefore reduce to

$$R^2=\frac{\sigma}{\rho}Q^2\cos\alpha,$$

$$G=R-\frac{\kappa Q^2}{2nh}\frac{\sigma}{\rho}\sin\alpha. \quad . \quad . \quad . \quad . \quad (43)$$

$$=R-\frac{\kappa R^2}{2nh}\tan\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

It will be noticed that k is not explicitly involved in these equations, so that the results obtained for shallow water do not depend on the eddy-viscosity, provided that this is such that the approximation is valid.

If $k=100$ cm.^{2}/sec., $n=10^{-5}$/sec., the condition for this is that $\frac{1}{2}h\sqrt{(n/k)}$ is small, which requires that h shall be small compared with 60 metres. Suppose now that $Q=10$ m./s., and that $\alpha=\frac{1}{4}\pi$. Then $R=60$ cm./sec. by (43), and the second term in (44) will be large compared with the first, provided that h is small compared with 60 metres. Hence the condition that we can approximate to (41) by neglecting positive powers of λ implies the condition that we can neglect the first term in (44), provided the wind exceeds 10 m./s., and we have therefore with sufficient accuracy

$$G=-\frac{\kappa Q^2}{2nh}\frac{\sigma}{\rho}\sin\alpha, \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

and

$$\frac{\partial\zeta}{\partial y}=-\frac{2nG}{g},$$

$$=\frac{\kappa Q^2}{gh}\frac{\sigma}{\rho}\sin\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

These results are strikingly different from the corresponding equations (34) and (39) for deep water. In fact, when α is zero, the wind produces no slope of the surface in the case just discussed. On the other hand, in deep water the maximum slope is produced by a wind parallel to the shore, and no slope is produced by a wind at right angles to the shore. In shallow water the maximum slope is produced by a wind at right angles to the shore, and no slope is produced by a wind parallel to the shore.

In the shallow water case equation (21) reduces to

$$W = \mathrm{R}e^{i\gamma} + \frac{1}{2}\frac{\kappa h}{k}\left(\mathrm{R}^2 e^{i\gamma} + \frac{\sigma}{\rho}\mathrm{Q}^2 e^{i\alpha}\right),$$

$$= \mathrm{R} + \frac{1}{2}\frac{\kappa h}{k}\mathrm{R}^2(2 + i\tan\alpha). \quad . \quad . \quad . \quad . \quad (47)$$

The eddy viscosity is therefore explicitly involved in the surface-drift, which takes place at an inclination to the shore intermediate between 0 and $\tan^{-1}(\frac{1}{2}\tan\alpha)$, approximating to the former value for light winds and to the latter for strong winds.

If we have $Q = 10$ m./s., $h = 10$ m., and $\alpha = \frac{1}{4}\pi$, (46) gives

$$\frac{\partial\zeta}{\partial y} = -\frac{1}{6}\frac{\text{cm.}}{\text{km.}},$$

and if $k = 100$ cm.²/sec.,

$$W = (48 + 9i) \text{ cm./sec.}$$

Thus the surface-drift has a resultant velocity of about 49 cm./sec., inclined at 10° 30′ to the shore.

It may be noticed that if the friction of the wind on the water surface is redistributed uniformly through the depth of the water, the force per unit mass in the direction of y is $\dfrac{\kappa\sigma \mathrm{Q}^2 \sin\alpha}{\rho h}$, and the equation of motion at right angles to the coast, internal and bottom friction being neglected, is

$$\frac{\partial v}{\partial t} + 2nu = \frac{\kappa\sigma \mathrm{Q}^2 \sin\alpha}{\rho h} - \frac{g}{\rho}\frac{\partial\zeta}{\partial y}.$$

When we give $\dfrac{\partial\zeta}{\partial y}$ the value found in (41), the right side vanishes. Hence the slope of the surface in the case of shallow water is the same as that of an equilibrium tide unaffected by internal and bottom friction and with the skin friction at the surface uniformly distributed through the depth.

The time needed to establish the steady state.

Let us consider the conditions in the open ocean, with no complication arising from the slope of the free surface. The force per unit area acting across the surface is $\kappa\sigma Q^2$ in the direction of the wind. When a steady state is attained, which will ordinarily take about a day, the momentum relative to the solid earth of a vertical column of water of unit cross-section will therefore be $\kappa\sigma Q^2/2n$, at right angles to the wind. Accordingly, if we consider a surface inclined at an angle α to the wind, the volume of water crossing it in unit time will be $\kappa\sigma Q^2 \cos\alpha/2n\rho$ per unit length.

Now suppose that parallel to such a surface, at a distance l, a solid vertical barrier is suddenly inserted. The water will temporarily cross the surface considered at the same rate as before, but clearly this cannot be maintained, for in time enough water will have crossed to raise or lower the slope of the surface to the value found in the case of a steady motion obstructed by a shore, and further accumulation will not take place. To raise the slope to the equilibrium value $\frac{\partial\zeta}{\partial y}$, the amount of water required is evidently $\frac{1}{2}l^2\frac{\partial\zeta}{\partial y}$ per unit length. Hence the time taken to establish a steady slope within a distance l of the coast must be of order

$$\frac{n\rho l^2}{\kappa\sigma Q\cos\alpha}\frac{\partial\zeta}{\partial y} \quad\text{or}\quad \frac{2n^2\rho l^2 G}{\kappa g\sigma Q^2\cos\alpha}, \quad . \quad . \quad \text{by (9)}$$

In the case of a hurricane blowing over deep water, this reduces, by (34), to $\frac{n^{\frac{3}{2}}l^2}{gk^{\frac{1}{2}}}$. If $l=2\times10^7$ cm. $=200$ km., $n=4\times10^{-5}$/sec., $g=980$ cm./sec.^{2}, and $k=100$ cm.^{2}/sec., this is practically 10^4 secs. or 3 hours. Accordingly the steady conditions will be established within 200 km. of the shore in a few hours.

In the case of a light wind over deep water, the time required reduces, by (39), to $\frac{2n^2l^2}{\kappa g G}$. With our previous data and $G=18$ cm./sec., as on p. 121, paragraph (*b*), this is about 12 hours.

In the shallow water case, equation (45) gives the time as $nl^2\tan\alpha/gh$. With the previous data and $h=10$ metres, this becomes 16,000 secs., or $4\frac{1}{2}$ hours.

The time needed to convert open ocean conditions into a state of steady motion obstructed by a shore is therefore a fraction of a day for places within 200 km. of the coast. The time required to establish such conditions at greater distances from the shore is proportional to the square of the

distance. In general a steady motion in a confined sea will be established in a day from the commencement of the wind.

The motion of water in the open ocean under the influence of a cyclone or a large seasonal depression or elevation may be discussed on these lines. The drift is at right angles to the wind, towards the high pressure side, and therefore tends to lower the water-surface within the cyclone. (The direct effect of pressure on the water-surface is, of course, in the opposite sense, but is not part of the subject-matter of this paper.) The argument of the papers quoted on p. 114 applies to cases where the surface has no slope, and therefore only to the early stages of the disturbances. The conditions are, in general, those of light winds over deep water, and we see that the surface-current will be at 45° to the wind in the early stages, becoming more nearly along the isobars in the latter stages. For a cyclone with light winds and with horizontal dimensions large compared with 400 km., the time taken for the change will be of order 3 or 4 days; for a hurricane we have already seen that it is a few hours; and for a disturbance whose horizontal dimensions are large compared with 4000 km., the time is a few months. Hence in seasonal disturbances in the Pacific and Atlantic Oceans the surface-current should be at approximately 45° to the wind.

XI. *The Stress System of the Four-Dimensional Electromagnetic Field.* By S. R. MILNER, *D.Sc., F.R.S., Professor of Physics, The University, Sheffield* *.

IN certain respects, as has been shown in a recent paper †, the general electromagnetic field, when it is considered as a four-dimensional entity, appears as the exact and natural extension to four dimensions of the three-dimensional electrostatic field. In the latter the properties are expressed in terms of a vector function of position $\mathbf{e}$ "acting" in a certain line, say x, the orientation of which, along with that of the plane yz, to which it is perpendicular, is fixed for each point. In the electromagnetic field in four dimensions the properties may be expressed by a vector function of position

$$R=\{(\mathbf{e}^2-\mathbf{h}^2)^2+4(\mathbf{e}\,\mathbf{h})^2\}^{1/4}$$

"acting" in a certain plane, say xt, the orientation of which, along with that of the plane yz to which it is perpendicular, is fixed for each point. In the one case the fundamental electrostatic equations are equivalent to the statement that the vector $\mathbf{e}$ can be represented by lines of force, or in

* Communicated by the Author.

† Milner, Phil. Mag. xliv. p. 705 (1922).

other words that the flux of **e** over the cross-section of a Faraday tube is constant throughout its length. In the other the fundamental electromagnetic equations are equivalent to the same statement with respect to the vector **R**.

There is of course a difference between the two fields which is obvious to the senses and which is due to the fact that geometry in the t direction is different in character from the geometry of the xyz space. By using imaginary time $l=it$ the difference of the geometries disappears and the equivalence of the two fields is rendered complete, at any rate in these respects.

The stress system of a four-dimensional electromagnetic field forms another respect in which a similar exact equivalence holds with an electrostatic field in three dimensions. Every element of the latter may be regarded as being in equilibrium under the action of a tension of magnitude $\frac{1}{2}e^2$ along the lines of force combined with an equal pressure at right angles to them. This stress system may be represented with arbitrary axes x, y, z, as a tensor

$$\left.\begin{aligned} P_{xx}&=\tfrac{1}{2}(-e_x^2+e_y^2+e_z^2), & P_{xy}&=P_{yx}=-e_xe_y,\\ P_{yy}&=\tfrac{1}{2}(-e_y^2+e_z^2+e_x^2), & P_{yz}&=P_{xy}=-e_ye_z,\\ P_{zz}&=\tfrac{1}{2}(-e_z^2+e_x^2+e_y^2), & P_{zx}&=P_{xz}=-e_ze_x.\end{aligned}\right\} \quad . \quad . \quad (1)$$

Here P_x represents the vector stress on a unit area of the plane yz whose normal is the x axis, and P_{xx}, P_{xy}, P_{xz} are the components of P_x along x, y, z. To see the equivalence of (1) with the stress system as described, choose the axes so that

$$e_x=e, \qquad e_y=e_z=0.$$

We get

$$P_{xx}=-\tfrac{1}{2}e^2, \qquad P_{yy}=P_{zz}=+\tfrac{1}{2}e^2, \qquad P_{xy}=0, \text{ etc.}$$

The components along the axes of F, the resultant force per unit volume acting on an element $dxdydz$ of the field, are given by

$$F_x=\frac{\partial P_{xx}}{\partial x}+\frac{\partial P_{yx}}{\partial y}+\frac{\partial P_{zx}}{\partial z}, \text{ etc.}$$

The equilibrium of the stress follows from the fact that, as a result of the fundamental electrostatic equations (curl **e** $=0$, div **e** $=0$) each component of F reduces to zero.

Consider now in the general electromagnetic field in four dimensions the following tensor, which has been given by Sommerfeld (*Ann. d. Phys.* xxxii. p. 769, 1910) as the most general symmetrical tensor product determinable from the two six-vectors of the electromagnetic field (**h**, $-i$**e**), ($-i$**e**, **h**).

$$\left.\begin{aligned}
P_{xx}&=\tfrac{1}{2}(-e_x^2+e_y^2+e_z^2-h_x^2+h_y^2+h_z^2),\\
P_{yy}&=\tfrac{1}{2}(-e_y^2+e_z^2+e_x^2-h_y^2+h_z^2+h_x^2),\\
P_{zz}&=\tfrac{1}{2}(-e_z^2+e_x^2+e_y^2-h_z^2+h_x^2+h_y^2),\\
P_{ll}&=\tfrac{1}{2}(-e_x^2-e_y^2-e_z^2-h_x^2-h_y^2-h_z^2);\\
P_{xy}&=P_{yx}=-\ (e_x e_y+h_x h_y),\\
P_{yz}&=P_{zy}=-\ (e_y e_z+h_y h_z),\\
P_{zx}&=P_{xz}=-\ (e_z e_x+h_z h_x);\\
P_{xl}&=P_{lx}=-i(e_z h_y-e_y h_z),\\
P_{yl}&=P_{ly}=-i(e_x h_z-e_z h_x),\\
P_{zl}&=P_{lz}=-i(e_y h_x-e_x h_y).
\end{aligned}\right\} \quad . \quad . \quad (2)$$

This may be interpreted as a four-dimensional stress, where P_x stands for the four vector stress on a unit volume of the hyperplane yzl whose normal is the x axis, and P_{xx}, P_{xy}, P_{xz}, P_{xl} are the components of P_x along x, y, z, l. The components along the four axes of the resultant force per unit hypervolume on an element $dxdydzdl$ are then given by the equations

$$\begin{aligned}
F_x&=\frac{\partial P_{xx}}{\partial x}+\frac{\partial P_{yx}}{\partial y}+\frac{\partial P_{zx}}{\partial z}+\frac{\partial P_{lx}}{\partial l},\\
F_y&=\frac{\partial P_{xy}}{\partial x}+\frac{\partial P_{yy}}{\partial y}+\frac{\partial P_{zy}}{\partial z}+\frac{\partial P_{ly}}{\partial l},\\
F_z&=\frac{\partial P_{xz}}{\partial x}+\frac{\partial P_{yz}}{\partial y}+\frac{\partial P_{zz}}{\partial z}+\frac{\partial P_{lz}}{\partial l},\\
F_l&=\frac{\partial P_{xl}}{\partial x}+\frac{\partial P_{yl}}{\partial y}+\frac{\partial P_{zl}}{\partial z}+\frac{\partial P_{ll}}{\partial l}.
\end{aligned}$$

For simplicity we confine the discussion to regions where there is no electric charge. As a result of the fundamental electromagnetic equations each component of F reduces to zero, so the element is subject to no resultant force. There is no torque also, in consequence of the equality of P_{xy} with P_{yx}, etc. It follows that the general electromagnetic field in hyperspace with imaginary time may be looked upon as being in statical equilibrium under the action of the stress system (2)*.

* This is not put forward as a new result, but I have not seen it stated anywhere in this form. It is of course only by representing time as imaginary space that the complete symmetry of l with x, y, z appears. With real time the equations $F_x=0$, etc., are expressions of well-known theorems, the first three that the time-rate of increase of the momentum of a space element of the field is equal to the resultant of Maxwell's

To see the real meaning of the stress tensor (2), its expression can be simplified by changing the axes in precisely the same way as the three-dimensional stress tensor was simplified. The possibility of this follows from a proposition which was proved in the previous paper: "In the general electromagnetic field in four dimensions it is always possible to choose the axes of x, y, z, and l at any point in such directions that the electric and magnetic forces are collinear and along the direction of x*."

If we then choose the axes so that this condition is satisfied, *i. e.* so that

$$e_x=\mathrm{E},\ h_x=\mathrm{H},\ e_y=e_z=h_y=h_z=0,$$

the expression for the Sommerfeld tensor (2) becomes

$$\mathrm{P}_{xx}=\mathrm{P}_{ll}=-\tfrac{1}{2}(\mathrm{E}^2+\mathrm{H}^2)=-\tfrac{1}{2}\mathrm{R}^2,$$
$$\mathrm{P}_{yy}=\mathrm{P}_{zz}=+\tfrac{1}{2}(\mathrm{E}^2+\mathrm{H}^2)=+\tfrac{1}{2}\mathrm{R}^2,$$

with all the remaining components zero.

The tensor (2) consequently represents a stress system consisting of a tension of $\frac{1}{2}\mathrm{R}^2$ in all directions in the xl plane of these axes, combined with an equal pressure in all directions of the absolutely orthogonal yz plane. This is an exact extension to four dimensions of the electrostatic stress system in three. The five-vector $(\mathrm{R}, i\mathrm{R})$ (six-vector with its two parts equal) consists of R associated with the yz plane by "acting" along every direction in the xl plane perpendicular to yz, combined with $i\mathrm{R}$ similarly associated with the xl plane. It is the natural extension to four dimensions of the vector **e** in the three-dimensional electrostatic system. In the four-dimensional system half the square of the vector in each case gives the stress over the corresponding plane (*i. e.* $\frac{1}{2}\mathrm{R}^2$ over yz, and $\frac{1}{2}(i\mathrm{R})^2$ over xl). There is here a complete symmetry of the tensions and pressures, as against only a partial symmetry in the three-dimensional case.

electromagnetic stress on it, and the fourth that the rate of increase of the energy density at a point is equal to the convergence there of Poynting's energy flux. But even with real time the equations may also be looked on as the expression of the static equilibrium of a four-dimensional system of stress, the formulæ being direct extensions of those of the three-dimensional system when allowance is made for the different geometry of hyperspace in the t direction. The result is consequently implicitly contained in the formulæ for the tensor in the general relativity theory. (See Eddington, 'Mathematical Theory of Relativity,' p. 182 (1923), where the general properties of the tensor are discussed.)

* Choosing the l-axis is equivalent to transforming the field for an arbitrary velocity of the observer. To produce collinearity of **e** and **h**, strictly speaking only the absolutely orthogonal planes xl and yz need be fixed.

XII. *The Emission of Secondary Electrons from Metals under Electronic Bombardment.* By Frank Horton, *Sc.D., F.R.S.*, and Ann Catherine Davies, *D.Sc.**

In the Philosophical Magazine for May 1923, Mr. E. W. B. Gill has criticised certain statements made by the writers in a paper describing the results of an investigation of the effects of electron collisions with platinum and with hydrogen †. The main object of this investigation was to test whether the secondary emission of electrons obtained by the electronic bombardment of platinum could be attributed to ionization of hydrogen occluded in the surface of the metal. In the introduction to the paper, the summary given by Campbell ‡ of the work of Lenard, Baeyer, Gehrts, etc., and of Campbell himself, is referred to, and the views then prevailing with regard to reflexion and secondary emission are stated. As a preliminary to our main investigation, a few series of observations were taken of the current to the bombarded platinum plate with gradually increasing values of the velocity of impact of the primary electrons, using a constant difference of potential between the glowing filament which was the source of the primary electrons and the grid on which the electrons leaving the plate were collected. A curve representing a typical series of observations is given in the paper, and the various parts of this curve are there interpreted in accordance with the conclusions established by other workers.

Mr. Gill has performed some experiments with an apparatus differing in several important respects from that employed by us, and the results of these experiments have led him to criticise the interpretation which we put upon the various parts of the curve already referred to. These criticisms do not affect the main conclusions of our investigation into the origin of the ionization which occurs at the platinum surface under the electronic bombardment, for the proof of the occurrence of this ionization was the detection, under suitable conditions, of positive ions leaving the plate. As, however, we are unable to admit the applicability of Mr. Gill's contentions to experiments made with the apparatus and arrangements employed by us, it is perhaps desirable that his statements should not be allowed to pass unanswered.

The criticisms fall into two main divisions:—

(*a*) Those based on the argument that experiments such

* Communicated by the Authors.

† F. Horton and A. C. Davies, Proc. Roy. Soc. A. vol. xcvii. p. 23 (1920)

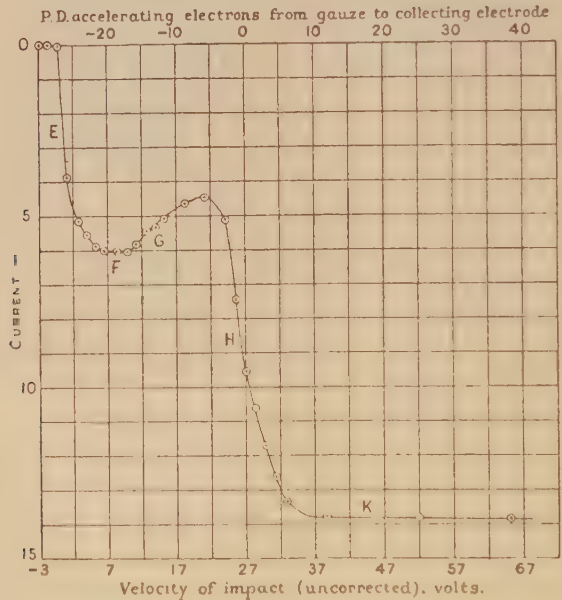
‡ N. R. Campbell, Phil. Mag. vol. xxv. p. 803 (1913).

as those of which the results are represented by our plate current–velocity of impact curve are incapable of discriminating between reflexion and secondary emission, or of enabling the critical voltages for the beginning of these effects to be determined with accuracy.

(*b*) Those based on the argument that a considerable error is introduced by assuming that the electric fields on either side of the grid are independent of each other.

With regard to (*a*) the authors did not claim that the experiments criticised did distinguish between the reflexion

Fig. 1.



Curve representing the variation of the current measured by the electrometer connected to the plate, as the velocity of impact of the primary electron stream is gradually raised. [Reduced from the figure in the Proceedings of the Royal Society, vol. xcvii.]

of the primary electrons and a genuine secondary emission, but merely that their results could be logically interpreted in accordance with the views current at the time. They do, however, still hold the view that the sharpness of the bend between the parts F and G of the curve in question (which, for convenience of reference, is reproduced here) indicates the commencement of a new effect when the velocity of impact reaches the value which it has at the point of inflexion.

The sharpness of this bend is not shown in the inaccurate reproduction of the curve given in Mr. Gill's paper. That the occurrence of an inflexion at this particular velocity of impact is significant, and not, as Mr. Gill suggests, fortuitous, is supported by the fact that with a suitable arrangement of electric fields, positive ions could first be detected leaving the bombarded plate when the velocity of impact of the primary electrons reached this particular value. The conclusion that there is a definite velocity of impact at which the true secondary emission of electrons from a bombarded metal surface begins, is supported by the results of the recent experiments of H. E. Farnsworth * on the electronic bombardment of nickel. In these experiments the electrons leaving the bombarded nickel plate were collected on an independent conductor, and it was found that at about 9 volts velocity of impact the ratio of the current leaving the nickel plate to the primary current began to increase rapidly. Moreover, by separate experiments it was shown that the distribution of velocities among the electrons leaving the plate changed abruptly at this same velocity of impact in the manner to be expected if a genuine secondary emission were beginning.

With regard to the criticism (*b*) :—Mr. Gill's experiments were made with a three-electrode Marconi valve, which does not appear to the authors to have been suitable for investigating the emission of secondary electrons from a metal under electronic bombardment. The grid and plate of the valve used were coaxial cylinders with the filament as axis. The grid had a square mesh of side about 1·5 mm. and the open ends were 1 cm. in diameter, the diameter of the cylindrical anode being 2·5 cm. The effects of the size of the grid mesh and of the open ends of a cylindrical grid occupied the attention of the authors in connexion with the design of a thermionic valve in the early days of the war, and they agree with Mr. Gill that, with such an arrangement as he employed, when the potential of the grid (V_g) is kept constant and that of the plate is increased, there may be a marked increase in the proportion of the electrons leaving the filament which reach the plate owing to the imperfect shielding action of the grid.

The apparatus used by the authors was constructed so as to avoid errors due to incomplete screening of the various electric fields from each other. The filament was enclosed in a platinum cylinder through a small hole in the centre of the top of which the electrons which formed the primary

* H. E. Farnsworth, Proc. Nat. Acad. Sci. vol. viii. p. 251 (1922).

bombarding stream were emitted. The stream was prevented from spreading laterally (and so introducing complications) by means of a strong magnetic field at right angles to the electrodes. The platinum gauze which was used as the grid was of fine mesh, the spaces being 0·0526 sq. mm. in area (or only about one-fortieth of the area of the spaces in Mr. Gill's grid), and the width of the wires of the gauze used was about one-third of the distance separating them. Moreover, the gauze formed the base of a platinum cylinder which enclosed the platinum plate, the arrangement (a diagram of which is given in our paper) being such that no electrons from the filament could have reached the plate without passing through the gauze, even if the magnetic field had not been used to keep the stream central in the tube. With such an apparatus it is obvious that any defect in the screening action of the grid is very much smaller than in the case of the valve used by Mr. Gill.

Let us now consider Mr. Gill's criticism of our statement in regard to the large increase of negative current which is indicated by the part H of the curve reproduced. It will be seen that as the potential difference between the gauze and the plate (the collecting electrode) was gradually raised, the negative current to the plate began to increase while this electrode was still about 5 volts negative to the grid. This increase of negative current was attributed by the authors to a decrease in the number of secondary electrons leaving the bombarded plate rather than to an increase in the number of primary electrons reaching it. Mr. Gill accuses the authors of having evaded an explanation of this "difficulty," and makes the statement that, if the field between the grid and the plate were determined simply by the applied potentials, it would certainly remove the electrons to the grid as long as V_p was less than V_g. The authors were not aware that the phenomenon presented any difficulty of interpretation, for it seemed to them not at all improbable that a difference of potential of a few volts would be required between the plate and the grid in order to cause the current due to the secondary electrons liberated at the plate to attain its saturation value. The surface of the plate is not perfectly smooth, and the holes in it (of molecular dimensions) serve to screen many of the emitted electrons from the action of the field, so that a potential difference (of about 5 volts in the curve reproduced) was required to drive all of these electrons to the grid.

Mr. Gill seeks to explain the phenomenon in another way, as being due to the field near our plate becoming zero and

reversing before the value of V_p reaches that of V_g (presumably 5 volts before this equality is established), and he refers to Maxwell's investigation of the theory of a grating of parallel wires *. Dr. Appleton, of the Cavendish Laboratory, who has made an extensive investigation of the screening action of grids, observing this fallacy in Mr. Gill's interpretation, was good enough to inform the authors that in an apparatus of the form used by them the shielding must have been almost complete, and that the direction of the field near the plate could not have reversed until $(V_g - V_p)$ became a very small fraction of V_p. In Maxwell's investigation, to which Mr. Gill refers, the case considered is that of a grating of parallel wires between two parallel plane electrodes, the grating being thus similarly situated to the platinum gauze in our experiments. If one considers the wires in one direction only in our gauze and applies Maxwell's formula to the case, one obtains the result that the surface density of the electrification on the plate becomes zero when $(V_g - V_p)$ is about $\frac{1}{300}$ of V_p. The shielding of the actual grid is, of course, more nearly perfect than this because of the cross-wires, at right angles to the wires considered, which form another similar grating. It will thus be seen that in the series of observations represented by our curve, the difference of potential between the grid and the plate when the direction of the field near the plate changed sign cannot possibly have been as much as $\frac{1}{10}$ of a volt—the limit of accuracy claimed by us for the results of experiments of this nature. Thus the increase of negative current beginning while the grid was still about 5 volts positive to the plate cannot have been due to the cause suggested by Mr. Gill.

It may here be stated that the magnitude of the difference of potential between the grid and the plate at which the direction of the field reverses may be investigated experimentally by observations on the direction of the resultant photoelectric current with small applied differences of potential between these electrodes, while a strong constant field is maintained on the other side of the grid. Such an investigation had convinced the authors that any deficiency of the shielding action of the grids used by them was negligible in the circumstances in which these grids were employed.

The proof that the shielding action of the grid was almost perfect disposes of Mr. Gill's contention that, because the authors made no allowance for an increase, with increasing plate potential, of the primary current to the plate at the

* 'Electricity and Magnetism,' vol. i. 3rd edn. p. 312.

expense of the primary current to the grid, their estimate of about 9 volts as the equivalent velocity of emission of the fastest secondary electrons is unreliable, and its agreement with the results of Lenard and others accidental. It should perhaps be mentioned that Farnsworth, in the recent experiments already referred to, obtained evidence of the presence among the electrons leaving the bombarded plate of a small percentage of electrons having velocities nearly up to that of the primary stream. Farnsworth concluded that these electrons were reflected from the plate and did not form part of the genuine secondary emission. The presence of high speed reflected electrons could not, of course, be detected by the method of investigation used by the writers.

Since the publication of the authors' experiments on the emission of secondary electrons from platinum, the results of several similar investigations, in addition to that of Farnsworth, have appeared, notably the experiments of Millikan and Barber* (which Mr. Gill appears to have overlooked), and the more recent investigations of McAllister†. This series of researches, carried out in the Ryerson Physical Laboratory, has led to the conclusion that the maximum velocity of emission of the secondary electrons from copper under electronic bombardment is about 10 volts, and that a potential difference of a few volts is needed between the plate and the grid in order to obtain the saturation value of the current carried by the electrons leaving the plate. Moreover, this latter result is explained by Millikan and Barber as being due to many of the secondary electrons being liberated in "pockets" or holes in the plate, a view similar to that held by the authors.

Mr. Gill's curves are sufficient evidence of the complexity of the factors which need to be taken into consideration in attempting to interpret the results of an investigation of the emission of secondary electrons from a metal surface subjected to electronic bombardment, when an apparatus is used in which there may be a serious interpenetration of the electric fields. The authors make no attempt to discuss Mr. Gill's interpretation of his own results; they only desire to refute the suggestion that the defects which made it impossible for Mr. Gill to draw any useful conclusion with regard to critical voltages from his experiments were, as he assumes, inherent in their own investigation.

* R. A. Millikan and I. G. Barber, Proc. Nat. Acad. Sci. vol. vii. p. 13 (1921); and I. G. Barber, Phys. Rev. vol. xvii. p. 322 (1921).

† L. E. McAllister, Phys. Rev. vol. xxi. p. 122 (1923).

XIII. *The Relative Intensity of X-Ray Lines.* By Frank C. Hoyt, *Ph.D., National Research Fellow* *.

I. *Introduction.*

THE uniform success of Bohr's correspondence principle † as a criterion for the occurrence and state of polarization of spectral lines and as a means of at least qualitative estimation of relative intensities makes it of the greatest importance to attempt its more exact quantitative application. This principle only states the general method by which we are to obtain an estimate of the relative frequency of occurrence of different quantum transitions, but does not lead to an unambiguous expression for the intensity of a given line. Among various plausible expressions, however, we might hope to discriminate by a comparison with experimental values. Kramers ‡ has been able to obtain approximate values for the relative intensities of the fine-structure components of hydrogen and helium lines as well as of the components into which these lines are split up in the presence of electric fields. In this connexion he has suggested how more accurate values may be calculated, but the exact comparison with experiment is not possible for these problems at present, and it may be that the question may be more advantageously attacked in the X-ray region where, except for the increased complexity of the atomic model, the experimental and theoretical problem is in certain respects simpler.

This paper contains a brief discussion of the factors that must be taken into account in a complete theory of the intensity of X-ray lines, and some calculations are made on the basis of the simplest atomic model possible.

II. *Correspondence Principle and Intensity of Spectral Lines.*

The correspondence principle establishes a relation between the quantum theory of radiation and the motion of the particles in the atom, which shows the closest possible analogy with the classical relation between electromagnetic

* Communicated by Prof. N. Bohr.

† N. Bohr, "On the Quantum Theory of Spectral Lines," *D. Kgl. Danske Vidensk. Selsk. Skrifter*, 1, iv.₁ (1918). Compare also 'Three Essays on the Theory of Spectra and Atomic Constitution,' Cambr. Univ. Press, 1922.

‡ H. A. Kramers, "Intensities of Spectral Lines," *D. Kgl. Danske Vidensk. Selsk. Skrifter*, 8, iii.₃ (1919).

radiation and the motion of a system of electrified particles. The principle has been developed by the consideration of systems for which the solution of the mechanical equations of motion can be represented as a superposition of harmonic vibration components *. This means that the displacement of every particle can be represented as a function of the time t by a trigonometric series of the type

$$\rho=\sum_{-\infty}^{+\infty}\ldots\sum_{-\infty}^{+\infty} C_{\tau_1\ldots\tau_s}\cos\left[2\pi(\tau_1\omega_1+\ldots\tau_s\omega_s)t+\delta_{\tau_1\ldots\tau_s}\right] \quad . \quad . \quad (1)$$

where the quantities $\omega_1\ldots\omega_s$ are the so-called fundamental frequencies of the motion, while $\tau_1\ldots\tau_s$ are integers. Since obviously a similar expression will hold for the component in a given direction of the resultant electric moment of the atom, the radiation which on ordinary electrodynamics would be emitted by the system would at any moment be composed of trains of harmonic waves with frequencies

$$\tau_1\omega_1+\ldots\tau_s\omega_s. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

According to the quantum theory the stationary states are defined by a set of conditions which may be written

$$I_1=n_1h\ldots I_s=n_sh, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $n_1\ldots n_s$ are integers and h Planck's constant, while $I_1\ldots I_s$ are quantities describing certain mechanical properties of the motion. The energy of the system depends on the quantities $I_1\ldots I_s$ in such a way that the energy difference δE of two neighbouring mechanical motions is expressed by

$$\delta E=\omega_1\delta I_1+\ldots\omega_s\delta I_s. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The frequency emitted during a transition between two stationary states for which $n_1\ldots n_s$ is equal to $n_1'\ldots n_s'$ and $n_1''\ldots n_s''$ respectively, is given by the well-known relation

$$\nu=\frac{1}{h}\left[E(n_1',\ldots n_s')-E(n_1''\ldots n_s'')\right]. \quad . \quad . \quad (5)$$

According to (5) we consequently have

$$\nu=\int_{'}^{''}\omega_1\delta I_1+\ldots\omega_s\delta I_s. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

If for the "path of integration" in the s-dimensional "$I_1\ldots I_s$ space" we chose a straight line connecting the

* A brief survey of the principles on which the applications of the quantum theory to atomic problems are based will be found in an article by N. Bohr, *Zs. für Phys.* xiii. p. 113 (1923).

points corresponding to the stationary states, that is, if we put

$$\left.\begin{aligned} I_1 &= h[n_1'' + (n_1' - n_1'')\lambda], \\ I_2 &= h[n_s'' + (n_s' - n_s'')\lambda], \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where λ is a quantity which takes all values from 0 to 1, the expression (6) assumes the simple form

$$\nu = \int_0^1 [(n_1' - n_1'')\omega_1 + \ldots (n_s' - n_s'')\omega_s] d\lambda, \quad . \quad (8)$$

and states that the frequency of the emitted radiation may be considered as a mean value of the frequency of one of the harmonic components occurring in the motion of the system, represented by (1), if we put

$$\tau_1 = n_1' - n_1'', \ldots \tau_s = n_s' - n_s''. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The correspondence principle states now that the presence in the electrical moment of the system of the harmonic component of this frequency must be regarded as the cause of the occurrence of the transition considered. Taking the view that the transitions are spontaneous processes *, this principle leads us to assume that the probability that a certain transition takes place within a given time-interval may be estimated from the amplitude of the corresponding harmonic component by a comparison with the intensity of the radiation to which on classical electrodynamics this harmonic component would give rise. On ordinary electrodynamics the amount of radiation emitted in unit time from an electron performing linear harmonic vibrations with frequency ω and amplitude C is given by

$$\frac{\Delta R}{\Delta t} = (2\pi)^4 \frac{2e^2}{3c^3} C^2 \omega^4 ; \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

and we may thus hope to effect such a comparison if we estimate the probability of transition by a process of averaging in such a way that, if we write

$$A'_{,,} h\nu = (2\pi)^4 \frac{2e^2}{3c^3} Q^2 \nu^4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where $A'_{,,}$ denotes the probability of transition in unit time, $Q^2\nu^4$ can be considered as a kind of mean value of $C^2\omega^4$ where ω and C are the frequency and amplitude of the corresponding harmonic component. We know that ν may be represented as an average value of the ω's by means of equation (8), but in the present state of the quantum theory

* *Cf.* A. Einstein, *Phys. Zeit.* xviii. p. 121 (1917).

we are left to conjectures in regard to the calculation of Q. However, as Kramers has already suggested, a reasonable procedure would be to take Q as an average of C over the same path in the I-space as may be used for ν. Thus we may immediately write down as possibilities

$$Q=\int_0^1 C_\lambda d\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\,a)$$

and

$$Q^2=\int_0^1 C_\lambda^2 d\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\,b)$$

or, since Q and ν are to be averaged along the same path, we may also write

$$Q\nu^2=\int_0^1 C_\lambda\omega_\lambda^2 d\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\,c)$$

and

$$Q^2\nu^4=\int_0^1 C_\lambda^2\omega_\lambda^4 d\lambda. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12\,d)$$

Considering our present ignorance of the mechanism of the emission of spectral lines, it seems hardly possible to distinguish between these by *a priori* arguments, and yet the differences between them are not altogether without theoretical significance. Thus (12*d*) is to be preferred if we wish to average the rate of emission of energy, as given by ordinary electrodynamics, during a slow change from one stationary state to another. In such a case, of course, the radiation would not be monochromatic; and if we wish rather to emphasize the fact that in reality it is, we may make the comparison with a monochromatic oscillator of which the amplitude only is found by an averaging process. We are then led to calculate Q by (12*a*) or (12*b*) and to take ν as the actual frequency. We have further to consider Kramers' * suggestion of a logarithmic method of averaging. In this way we obviate the choice between taking the average of a square or squaring the average of the first power, since both give the same result. Thus we can write also

$$\log Q^2=\int_0^1 \log C_\lambda^2 d\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad (13\,a)$$

or

$$\log Q^2\nu^4=\int_0^1 \log C_\lambda^2\omega_\lambda^4 d\lambda. \quad . \quad . \quad . \quad . \quad (13\,b)$$

* *Cf.* H. A. Kramers, *loc. cit.*, footnote, p. 46. The general arguments brought forward there against the strict validity of any simple mean value of the type in (12) or (13) cannot be maintained. *Cf.* N. Bohr, *Zs. für Physik*, xiii. p. 149 (1923).

We have, then, at least six conceivably possible methods of calculating $Q^2\nu^4$. The problems considered by Kramers offered no sufficient basis for a distinction between them by means of comparison with experiment; and it is the purpose of this paper to examine the way in which this may be done by means of measurements on the relative intensity of X-ray lines.

III. *X-ray Spectra and the Correspondence Principle.*

According to the quantum-theory picture, the excitation of characteristic X-rays consists in the removal of an electron from one of the inner orbits, and the atom is then ready for emission, which takes place by transition of an electron from another orbit to fill the place left vacant. The intensity of the radiation emitted depends on the relative probability or frequency of occurrence of this transition as compared with other possible ones. Denoting, as above, the probability that a spontaneous transition from a state (′) to a state (″) takes place within a time-interval Δt, by $A'_{''}\,\Delta t$ we get for the total amount of radiation R emitted in unit time as a consequence of a transition of this type

$$R = N'A'_{''}\,h\nu,$$

where N' is the mean number of atoms present at any moment in the state (′). From this expression it follows, in the first place, that the relative intensity of the X-ray lines which correspond to the various transitions which may occur after the removal of the same electron in the atom, and which have the same critical potential of excitation, are independent of the conditions under which the lines are produced *; and this conclusion is confirmed by the experiments of Webster † and Wooten ‡.

The theoretical estimation, by the correspondence principle, of the relative intensities of such lines requires of course a definite picture of atomic constitution. The general outlines of such a picture are furnished by Bohr's § recent theory, in which the nature and arrangement of the electron orbits is consistent not only with the general character of the spectra

* *Cf.* S. Rosseland, Phil. Mag. lxv. p. 65 (1923).
† *Cf.* Webster, Phys. Review, vii. p. 599 (1916); ix. p. 220 (1917).
‡ *Cf.* Wooten, Phys. Review, xiii. p. 71 (1919).
§ N. Bohr, *Fysisk Tidsskrift*, xix. p. 153 (1921): in German translation, *Zs. für Phys.* ix. p. 1 (1922); in English translation, 3rd essay in the volume cited on p. 135, note.

of the elements, but also with the periodicity of their physical and chemical properties. The classification of the electron orbits in this theory is based on the assumption that the motion of each individual electron in the atom to a first approximation may be described as a plane periodic orbit, on which is superposed a uniform rotation in its plane. Thus each orbit is represented by a symbol n_k, where the "principal" quantum integer n is analogous to the quantum number which defines the stationary states of a periodic Keplerian motion, while the subordinate quantum number k corresponds to the azimuthal quantum integer appearing in Sommerfeld's theory for the fine-structure of the hydrogen lines. The radial quantum integer appearing in the latter theory is in our notation equal to n-k. For electron orbits well in the interior of the heavier atoms, the dimensions differ but little from those of orbits with the same quantum number described by an electron revolving about the nucleus in the absence of other electrons. For the outer parts of the atom, however, the form and dimensions of the orbits will differ considerably from those of such orbits, since here the effect of repulsion from the other electrons is no longer small compared with the attraction towards the nucleus.

The principal quantum integer determines to a first approximation the energy of the orbit, and we thus have a natural division into groups of orbits with the same value of n but different values of k. These groups correspond to the K, L, M, ... "levels" in X-ray spectra, and their further divisions into sub-groups is consistent with the classification of the X-ray energy levels worked out by Coster * and Wentzel †. The interpretation of these levels, and in particular their characteristic changes with atomic number, have been more fully discussed in their relation to this theory in a recent paper by Bohr and Coster ‡. It has also been possible to establish certain simple empirical combination rules which suggest the way in which the correspondence principle is to be applied here. In fact, it appears from the work of the last-named authors that we may distinguish from among the others certain so-called "regular" levels, which are purely of the n_k type. The number of these levels corresponds exactly to the number of sub-groups of electron orbits required by the theory, and the empirical rules for the

* Phil. Mag. xliii. p. 1070, xliv. p. 546 (1922).
† *Zs. für Phys.* vi. p. 84 (1921).
‡ *Zs. für Phys.* xii. p. 342 (1923).

occurrence of transitions between them are such as would be predicted by the application of the correspondence principle to central motion. The remaining levels may be called "irregular"; and although their detailed interpretation is still doubtful, a preliminary designation by means of two k numbers has been employed, because presumably they are closely connected with the interaction of *two* sub-groups resulting from the removal of an electron from one of them.

This representation of the X-ray levels in relation to the structure of electron orbits provides us with a basis for the quantitative application of the correspondence principle. In the case of transitions between regular levels, it seems not unreasonable to suppose that we can estimate the intensities by direct comparison with central orbits. In general it is not at present possible to determine exactly the motion in the electron orbits, but for those well in the interior of the atom the motion will be approximately Keplerian. Thus for the energy levels corresponding to the removal of an electron from one of these inner orbits we may make use of the work of Kramers on the fine-structure intensity of hydrogen and helium lines. We will, in the first instance, make our calculations as though the nuclear charge were constant, and as though we had to do with the probability of transition of only one electron. We can then consider how these results may be modified to take account of the fact that the effective nuclear charge is different for the different orbits, and that in a sub-group we have really several orbits with the same values of n and k.

IV. *Calculations and Results.*

For central motion of a single electron about the nucleus, the expression corresponding to (1) for the displacement in any direction must take the form *

$$\xi = \sum_{-\infty}^{+\infty} C_\tau \cos 2\pi(\tau\omega + \sigma)t. \quad . \quad . \quad . \quad . \quad (14)$$

Here ω and σ are the frequencies of oscillation of r and ϕ respectively. The amplitude coefficients depend only on τ,

* *Cf.* N. Bohr, *D. Kgl. Danske Vidensk. Selsk. Skrifter*, 8, iv. (1918), part i. p. 33. *Cf.* also E. P. Adams, 'The Quantum Theory,' Bull. Nat. Research Council, vol. i. part 5, p. 367.

and Kramers has shown that for the relativity ellipses we have*

$$C_\tau = -\chi I^2 \frac{1}{2\tau}\left[(1+\epsilon')J_{\tau-1}(\tau\epsilon)-(1-\epsilon')J_{\tau+1}(\tau\epsilon)\right], \quad . \quad (15)$$

where

$$\chi = \frac{1}{4\pi^2 N e^2 m}, \qquad I = nh,$$

$$P = kh,$$

$$\epsilon' = \frac{P}{I}, \qquad \epsilon = \sqrt{1-\epsilon'^2}.$$

Here n and k are the principal and subordinate quantum integers, and N is the atomic number. $J_{\tau-1}$ and $J_{\tau+1}$ are Bessel functions of the order indicated by the subscript.

To calculate the value of $Q^2\nu^4$ by any one of the methods suggested in the expressions in (12) and (13), we must first express the C_τ and ω for any transition for which $n'-n''=\pm\tau$ and $k'-k''=\pm 1$ as a function of λ by means of the relations

$$I = h[n''+(n'-n'')\lambda],$$

$$P = h[k''\pm\lambda],$$

corresponding to (7). It will, in general, be impossible to evaluate directly the integrals thus obtained, but by graphing C_τ as a function of λ and measuring the area under the curve, sufficiently accurate values may be obtained. In the case of the logarithmic integral (13 a), which becomes infinite for $\lambda=0$ for certain transitions, an approximate expression for small values of λ was used which could be integrated directly. The quantity ω which occurs in (12 a) and (12 b) is approximately the frequency of revolution in the orbit, and may be readily calculated as a function of λ. For the other methods of estimating $Q^2\nu^4$, measured values of ν can be used.

The values of $Q^2\nu^4$ have been calculated in this way, assuming the nuclear charge in (15) constant, for the transitions between regular levels that may take place after the removal of an electron from a 1_1 orbit—that is, for the K-series transitions. For these lines the relative intensity will be given by the relative values of $Q^2\nu^4$, as seen from (11), in so far as we have to do only with the probability of transition of a single electron. But, as already mentioned, we have, for each line, to consider the probability of transition to a 1_1 orbit from a number of orbits with the same values of n and k. It is then easily seen that we must multiply our value of $Q^2\nu^4$ by the number of electrons in the sub-group n_k to which the initial orbit belongs, which may be denoted by r.

* H. A. Kramers, *loc. cit.* p. 298.

Relative Intensities of the K-Series Transitions of Rhodium.

$$\nu_{\alpha_1}^4 : \nu_{\beta_1}^4 : \nu_{\beta_2}^4 = 1 : 1{\cdot}587 : 1{\cdot}724.$$

Relative Values of $Q^2\nu^4r$.

Transition.	1. $Q^2\nu^4 = \nu^4\int_0^1 C_\lambda^2 d\lambda.$	2. $Q\nu^2 = \nu^2\int_0^1 C_\lambda d\lambda.$	3. $\log Q = \int_0^1 \log C_\lambda d\lambda.$	4. $Q^2\nu^4 = \int_0^1 C_\lambda^2 \omega_\lambda^4 d\lambda.$	5. $Q\nu^2 = \int_0^1 C_\lambda \omega_\lambda^2 d\lambda.$	6. $\log Q^2\nu^4 = \int_0^1 \log C_\lambda^2 \omega_\lambda^4 d\lambda.$	Experiment.
α_1 $2_2 \rightarrow 1_1$ ...	1	1	1	1	1	1	1·
β_1 $3_2 \rightarrow 1_1$ ...	·458	·409	·293	·28	·21	·126	·29
β_2 $4_2 \rightarrow 1_1$ ...	·332	·262	·105	·10	·075	·0075	·05

The relative intensities of the K-series lines thus computed in six different ways are given in the table. The values of ν^1 used are the empirical values for rhodium, and the number of electrons in the sub-groups is assumed to be that given by Bohr's theory for this element—*i. e.*, 4 in the 2_2 group, 6 in the 3_2, and 6 in the 4_2 *. For other regular transitions than those leading to α_1, β_1, and β_2 there is no corresponding amplitude in the motion because $k'-k''=\pm 1$. The experimental values for rhodium as estimated from Webster's and Wooten's results are also included. Webster's values are for rhodium only, and uncorrected for absorption in the glass of the X-ray tube, while those of Wooten are for molybdenum and palladium, corrected for absorption, but with α_1 and α_2 unseparated. By combining the two we may get a value for β_1 which is probably correct to within 2 or 3 per cent. In the case of β_2 the value can hardly be regarded as much more than an estimation of the order of magnitude. It is unfortunate that there are no experimental values for other elements.

As to the lines of the K-series that involve irregular levels, it should be noted that the only one of appreciable intensity is α_2, which is about $\frac{3}{4}$ as strong as α_1. The line β_3, which is of this class, is very weak, and the corresponding companion of β_2 has not yet been observed. Lines corresponding to other transitions have not been found.

As before mentioned, the nuclear charge has been assumed constant, and hence does not affect the calculated relative values. We are able, of course, to calculate the effective nuclear charge in the stationary states, but we do not know how to represent it as a function of λ for purposes of averaging. But since the amplitude coefficients are inversely proportional to the effective nuclear charge, we see that the general effect will be to increase the values for β_1 and β_2. We may, however, make a very rough estimate of the order of magnitude from the empirical screening constants. In the case of the averages as taken in columns 1 and 2 of the table, the effect may be as high as 30 per cent. for β_1 and 100 per cent. for β_2. For the logarithmic average and the last two in which ω is within the integral sign, the effect

* It is to be noted, however, that according to Bohr's theory of atomic structure, the 4_2 orbit in rhodium belongs to a sub-group which is still in a state of development. Thus for the heaviest elements there are 8 4_2 orbits, and for some of the lighter ones there are only 4. This implies corresponding changes in the intensity of β_2 which are much larger than those due to changes in the relative values of ν^1. Similar characteristic changes with atomic number are to be expected for other lines, and their examination would furnish an interesting test of the theory.

will be considerably smaller, because the way in which the amplitude changes with λ is such as to emphasize the values where λ is small and the correction nearly negligible. Here the effect is probably from 5 to 10 per cent. in the case of β_1 and 30 to 50 per cent. in the case of β_2, which latter is comparable with the experimental error. The corrections will, of course, be still smaller for elements of higher atomic number, but for these we lack any empirical values.

For the lines of the L- and M-series the rather lengthy computations of the average amplitudes have only been made in a few cases, as the uncertainties due to changes in the nuclear charge are greater, and there are no empirical values for comparison. However, it appears that here also the values will be of the right order of magnitude so far as can be seen from estimation of intensities from photographic plates. It is much to be hoped that in the future we may have some experimental values for L-series lines that are free from errors due to absorption in the glass of the X-ray tube.

V. *Conclusions.*

It would seem safe to conclude from the comparison with empirical values that it is possible by means of averages between the stationary states of the type employed to obtain a satisfactory representation of the intensities of X-ray transitions between regular energy levels. As to distinguishing between the different averages, we may only say with some certainty that those in columns 1, 2, and 6 may be discarded as differing too widely from the empirical values. Between the other three we cannot distinguish very sharply, although, strictly speaking, the best agreement is for the results in column 3, where the intensity is calculated by comparison with a monochromatic oscillator, the amplitude of which is found by a logarithmic average. It will be well, however, to reserve judgment on this point until we have data on the heavy elements.

Acknowledgment must be made to Professor N. Bohr and Dr. H. A. Kramers for constant help and criticism during the course of this work.

National Research Council of America Fellowship,
Universitetets Institut for teoretisk Fysik,
Copenhagen, Denmark. April 1923.

XIV. *Poisson's and Green's Theorems in Riemann's* n-*manifold both when it is space-like and when time-space-like.* By

Prof. ALEX. MCAULAY, *M.A., University of Tasmania**.

THE notations of this paper had been used in two other papers sent to the Phil. Mag. a little earlier than the present paper and were not again explained. As, however, these two papers have never appeared, it may now be said (May 1923) that the notations will be found in "Multenions and Differential Invariants," I., II., III. (Proc. Roy. Soc., A: vol. xcix. p. 292, vol. cii. p. 210, and some vol. still unknown).

§ 1. *Scope of the paper and enunciation of the two principal theorems.*—In this § 1 Poisson's theorem is enunciated for each of the two species of Riemann's n-manifold. In § 2 a minute explanation of the meaning of the enunciations is given. The opportunity is used to present several aspects of matters affecting the theorems, which are at the same time of general import. The proof follows in § 3, and the paper concludes with a treatment of Green's theorem for the case of the n-manifold.

In a space-like manifold—that is, one all of whose n-dimensions are real—Poisson's theorem becomes

$$\left.\begin{aligned} &\mathrm{V}_0\nabla(\bar{\eta}\nabla\textstyle\int^n\!\!\int h\bar{m}\,db) = \pi_n\bar{m}, \\ \text{where}\quad &h = (\text{distance})^{2-n}/(n-2), \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

except when $n=2$, in which case $h=\log$ (distance).

In a time-space-like manifold—that is, a manifold of which one dimension is time-like or imaginary and the remaining $n-1$ are space-like or real—the theorem becomes

$$\left.\begin{aligned} &\mathrm{V}_0\nabla(\bar{\eta}\nabla\textstyle\int^{n-1}\!\!\int h\bar{m}\mathrm{V}_0\iota\,d\alpha) = \pi_{n-1}\bar{m}, \\ \text{where}\quad & \\ &h = (2z+\mathrm{V}_0{}^2\iota\nabla z)^{\frac{3-n}{2}}/(n-3),\ z=\tfrac{1}{2}(\text{distance})^2, \end{aligned}\right\} \quad . \quad (2)$$

except when $n=3$, in which case

$$(n=3),\quad h=\tfrac{1}{2}\log(2z+\mathrm{V}_0{}^2\iota\nabla z). \quad . \quad . \quad . \quad . \quad (3)$$

It may be stated also that, putting

$$\nu = \mathrm{N}\nabla z = \mathrm{V}_1\iota\mathrm{V}_2\eta\iota\nabla z, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

h may alternatively be defined by

$$h = (-\mathrm{V}_0\nu\eta^{-1}\nu)^{\frac{3-n}{2}}/(n-3), \quad . \quad . \quad . \quad . \quad . \quad (5)$$

except when $n=3$ as before.

* Communicated by the Author.

It must be explained that

$$\pi_2 = 2\pi, \quad \pi_3 = 4\pi, \quad \pi_4 = 2\pi^2,$$

and generally π_n is the solid angle subtended at a point by the whole of an n-manifold when the manifold is space-like. [The similar solid angle in a time-space-like manifold, but subtended by the whole of the manifold within either light semi-cone, or subtended by the whole without the cone, is in both cases infinite.] If k is any positive integer, the following are the values of π_n [except for π_2, though the formula is easily modified to cover the case by writing $2^{k-1}(k-1)!$ for the denominator on the right of the first of the following equations]:

$$\left.\begin{aligned} \pi_{2k} &= (2\pi)^k/[2\,.\,4\ldots(2k-2)], \\ \pi_{2k+1} &= 2(2\pi)^k/[1\,.\,3\ldots(2k-1)]. \end{aligned}\right\} \quad . \quad . \quad . \quad (6)$$

Here is a table for π_n, calculated by 5-figure logarithms, so that last-figure accuracy is not guaranteed :—

n.	π_n.	n.	π_n.	n.	π_n.	n.	π_n.
2	6·2832	8	32·470	14	8·3898	20	·51614
3	12·566	9	29·687	15	5·7217	21	·29294
4	19·739	10	25·502	16	3·7653	22	·16215
5	26·319	11	20·725	17	2·3967	23	·08765
6	31·006	12	16·023	18	1·4786	24	·04631
7	33·074	13	11·838	19	·88582	25	·02394

π_n rapidly diminishes beyond the limits of the table; $\pi_{100} = 10^{-38}\,2{\cdot}3683$. Note that a maximum of about 33 is reached at $n=7$.

π_n is defined as the area of a small sphere, radius r, in an n-dimensional space, divided by r^{n-1}. Thus the area and volume of the sphere are $\pi_n r^{n-1}$, and $\pi_n r^n/n$ respectively, and the circumference and area of a circle of the same radius in the same space are $\pi_{n-1}r^{n-2}$ and $\pi_{n-1}r^{n-1}/(n-1)$ respectively. The volume of the sphere is generated by the motion of a circle of varying radius, which moves parallel to itself and whose radius is $r\cos\theta$ when it is at a distance $r\sin\theta$ from the centre of the sphere. Hence we have

$$\pi_n/n = (\pi_{n-1}/(n-1))\,.\,2\int_0^{\frac{1}{2}\pi}(\cos\theta)^n\,d\theta.$$

From this we further deduce that

$$\pi_{n+2}/\pi_n = 2\pi/n,$$

and (6) then readily follows.

§ 2. *The meaning of the theorems explained.*—Take ρ for

position vector, and η for the fundamental covariant vector linity, so that the scalar form is $V_0 d\rho\eta d\rho$ (though in a space-like manifold, except when consistency forbids, it is better to reverse the sign of η). Whenever a bar, as with $\bar{\eta}$ and $\bar{m}$ in (1) and (2) of § 1, is placed over a symbol it is understood that the symbol is an invariant, covariant, or contravariant quantity multiplied by k, where $k = |\eta|^{\frac{1}{2}}$ and $|\eta|$ stands for the determinant of η. Such a barred symbol is called a density. [Will somebody invent a better word?—this clashes confusingly with density as applied to matter.] Thus $\bar{\eta}$ is defined to mean the frequently recurring linity $k\eta^{-1}$. $\bar{m}$ is k multiplied by a scalar invariant function of position; it appears as the analogue below of material density when Poisson's theorem is taken in its strictly original sense.

(1), so obviously like the original Poisson theorem, serves to introduce (2). Let, in the case of (1), $r(=T\rho)$ be the distance of the point ρ from the origin O, whether of three-dimensional Euclidean space or of n-dimensional Riemannian. In the three-dimensional case, if we put $h = r^{-1}$ [see (1)] and form the volume integral $\iiint hmdb$ for any given volume, including the origin, Poisson's equation asserts that

$$4\pi\bar{m} = \nabla^2 \iiint h\bar{m}\,db = S\nabla(\nabla \iiint h\bar{m}\,db),$$

the $\bar{m}$ on the left meaning the value at the origin. In the general Riemannian case h becomes $r^{2-n}/(n-2)$, where, when we are using equiradial coordinates, r may still be defined as $T\rho$, but in general must be defined as distance (along a geodesic) between the element db and the origin. The operator $\nabla^2 = S\nabla\nabla$ becomes the invariant operator

$$V_0\nabla(\bar{\eta}\nabla[\ \ ])$$

(not that the result is an invariant, but that it is an invariant density), which for the future we shall write as $V^2\nabla\bar{\eta}\nabla$. The 4π becomes π_n, and the equation becomes

$$V_0\nabla\bar{\eta}\nabla \int^n \int h\bar{m}\,db = \pi_n\bar{m}.$$

For the value of a quantity at the origin, or more generally at the point belonging to the operator *outside the integral*, we shall use the suffix O whenever for clearness it seems desirable. The other end of the distance defining h is the point belonging to the element db *inside the integral*. Such a point we will denote when it seems desirable by I, and a function at I by the suffix I. Thus in the present

case, putting in O and I everywhere,

$$V_0 \nabla_0 \bar{\eta}_0 \nabla_0 \int^n \int h_{0\mathrm{I}} \bar{m}_\mathrm{I} db_\mathrm{I} = \pi_n \bar{m}_0,$$

or

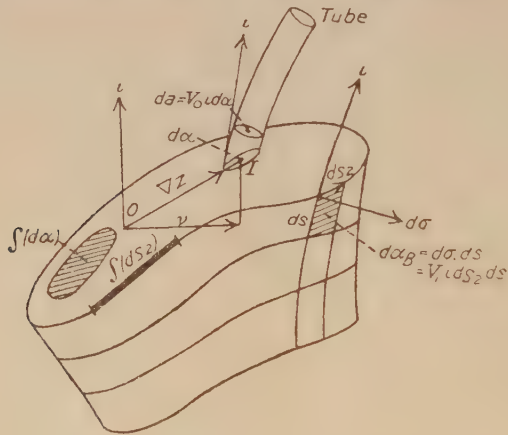
$$V_0 (\nabla \bar{\eta} \nabla)_0 \int^n \int h \bar{m}_\mathrm{I} db_\mathrm{I} \quad = \pi_n \bar{m}_0.$$

Even when we pass one or both of the symbols ∇, or the symbol η from outside to inside, we can retain the suffix O to remind us that the symbol still *belongs* to the point O, though it has temporarily migrated to a neighbour's location.

These fairly familiar matters have been detailed by way of introduction. Let me now try to indicate what I finally succeeded in satisfying myself of, in connexion with a time-space-like manifold. I started with the full expectation that, whether I could find it or not, there was some theorem very analogous to (1), and like (1) referring to an n-dimensional region, and treating all geodesics impartially *There is no such theorem.* Poisson's original theorem is interpretable in connexion with sources whose discharge takes place uniformly in all directions along geodesics. Now the solid angle (infinite) of all the three compartments (the inside of two semi-cones, past and future, and the spatial outside) into which the light-cone of a point partitions the manifold is in each case infinitely nearly concentrated on the cone. Hence, strengths of source being finite, the discharges when uniform in all directions, as prescribed, must be solely along the rays (*corresponding* to geodesics) of the cone. *But* there is nothing given by the fundamental form $V_0 d\rho\, \eta\, d\rho$ which enables us to divide the rays up into equal bundles: we cannot do so even in a Galilean manifold. The very conception of such a partition of the rays requires us to assign a time axis; it cannot be effected absolutely. This is our first great leading; we must suppose a congruence of time-like curves filling the manifold to be given, in order to fix the time axis everywhere—that is, if we would talk about sources discharging uniformly in all directions along geodesics (or their limiting rays on the cone). Proceeding to the details of the discharge so postulated, we naturally discover that the underlying phenomenon is $(n-1)$-dimensional, instead of n-dimensional. It is a phenomenon referring to an unbounded prescribed space-like surface (giving an $(n-1)$-dimensional *space*), at every point of which there is a prescribed contravariant time-like vector which may be regarded as specifying the *time* axis at that point.

The last sentence truly describes the mathematical requirements for the application of (2); a given wholly

space-like region of $n-1$ dimensions, $\int(d\alpha)$, wherein [see diagram drawn for $n=3$] is situated the point O, to which the outside operator belongs, and a given unit ($V_0\iota\eta\iota=1$)



Vectors: ι, ∇z, ν, $d\alpha$, d_{α_B}, $d\sigma$.

NOTE.—$ds\lambda_2$ is not a mere vector; it is a ${}_nV_2$ of $n-2$ dimensions. $d\alpha$ is a vector of $n-1$ dimensions. ds_2 is the common boundary of two intersecting regions $d\alpha$, $d\alpha_B$, and is expressible as a scalar multiple of $V_2 d\alpha d\alpha_B$.

contravariant vector ι at every point of the region. Nevertheless, it is better to consider the region as but one member of a family of the like, filling the manifold, and the contravariant vectors as the values, at the region selected, of the unit vectors specifying a whole congruence of curves throughout the manifold. The family and the congruence, one or both, or each partially, may be regarded as either prescribed or determinable at will according to the nature of the problem in hand. The space, as we may now call it, $\int(d\alpha)$ will sometimes below be taken to mean the light-cone, or rather any selected portion of it, but generally it is convenient to think of $\int(d\alpha)$ as a section of the manifold corresponding to a particular instant of coordinate time, however defined. We shall generally speak of it as the surface $t=$const., but when this would be misleading we can say instead $f=$const.

In (2) $d\alpha$ is an anti-density covariant vector (meaning a covariant vector multiplied by k^{-1}) element of the space or region. The curves ι through the boundary of the element $d\alpha$ form an elementary tube of which much will be said as our argument proceeds. The cross-section of the tube is $kV_0\iota d\alpha$. This in (2) is multiplied by $k^{-1}m$ as

well as by h. *It is this element* $V_0 \iota \, d\alpha$ *of tubular cross-section, not the corresponding portion of the region*, which is analogous to the element of volume in Poisson's original theorem; for $\bar{m}$ occurs outside the integral as well as inside, and h is the distance function connecting points O with points I. What still remains under the integral sign, namely $V_0 \iota \, d\alpha$, is analogous to what similarly remains in Poisson's case, the element of volume. This remark will be found true in physics generally. It is always the three-dimensional cross-section of a physical four-dimensional tube, whose generators are time-like curves, which is the mathematical representative of the physicist's three-dimensional volume. There seems no exception; certainly there is not in the domains of matter and electricity. [See my third R.S. paper on "Multenions and Differential Invariants." This paper was despatched many months before the meeting of the British Association last year (meaning the year 1921): I read that the tubes received some attention at that meeting.] It appears from the present paper that this dominance of the n-dimensional tubes is a mathematical consequence of the fundamental properties implied by η; it is not, on the contrary, something in addition imposed by an anthropomorphic legislator of Nature: the legislator has merely said, Let the world be Riemannian and for the rest you may do as you please.

The diagram implies correctly that da is a (scalar) component of the (vector) $d\alpha$ and that ν is a (vector) component of the (vector) ∇z, but is misleading as to the relative magnitudes of original and component of original. Thus actually

magnitude of $da >$ magnitude of $d\alpha$,

" $\nu >$ " ∇z.

The reader is warned to beware of misdirection from surprises of this kind, which, of course, are due to the imaginary dimension coming into comparison with the real dimensions.

As the diagram implies, it is often desirable to consider an n-dimensional region $\int (db)$ as bounded by two faces, that is, two members $\int (d\alpha)$ of the family, and by an edge consisting of generators $\int (\iota \, ds)$ of the tubes, ds being interval measured along ι. [db, $d\alpha$, $d\alpha_B$, $d\varsigma_2$ are all cases of $d\varsigma_a$, and are therefore all anti-density covariant multenions; db being a scalar, $d\alpha$ a vector, and $d\varsigma_2$ a ${}_nV_2$]. $d\varsigma_2$ is a common element of boundary of $\int (d\alpha)$ and $\int (d\alpha_B)$. $d\alpha_B$ may be taken as a parallelogram-like element of which $n-2$ sides belong to

$d\varsigma_2$ and the remaining one is ιds. It is convenient to introduce a vector $d\sigma$ parallel to $d\alpha_B$, but of the one dimension due to ds less than $d\alpha_B$. Thus

$$\left.\begin{aligned} d\alpha_B &= V_1\iota d\varsigma_2\,.\,ds = d\sigma\,.\,ds, \\ d\sigma &= V_1\iota d\varsigma_2, \\ V_3 d\varsigma_2 d\sigma = 0 &= V_3 d\varsigma_2 d\alpha_B = V_3 d\varsigma_2 d\alpha. \end{aligned}\right\} \quad . \quad . \quad . \quad (7)$$

[Note that between two consecutive members $\dot{\int}(d\alpha)$ the ds may or may not vary from tube to tube, ι, according to our prescription in the problem in hand.] $d\sigma$ is to be taken as the vector element of boundary of $\dot{\int}(d\alpha)$ when we explicitly use $da(V_0\iota d\alpha)$ as the element of integration of $\dot{\int}(d\alpha)$. da is the element of physical volume of the $(n-1)$-dimensional physical space $\int(d\alpha)$ and $d\sigma$ is the corresponding element of physical boundary; $d\sigma$ is *not in the region* $\dot{\int}(d\alpha)$, but instead it is normal to the tube ι.

Our original integration theorem ["Multenions and Differential Invariants," first paper, § 8, eq. (6); there are many misprints in the paper; db is omitted from the right of the equation here quoted] gives

$$\int^{n-1}\!\!\int \phi\, d\alpha = \int^{n}\!\!\int \phi_9 \nabla_9 db, \quad . \quad . \quad . \quad . \quad (9)$$

but in our present notation $d\alpha$ refers only to the faces, and the present form of (9) will be

$$\int^{n-1}\!\!\int \phi\, d\alpha_B = -\int^{n-1}\!\!\dot{\int} \phi\, d\alpha + \int^{n}\!\!\int \phi_9 \nabla_9 db. \quad . \quad (10)$$

If on bringing the two faces close together, and prescribing that ds is common to all the elements of tubes between the faces (this prescribes the consecutive member $\dot{\int}(d\alpha)$, when a particular one is given) we should have $db = V_0\iota d\alpha\,.\,ds$. If we could assume the two opposite ends of a tube to cause the sum $\phi\, d\alpha + \phi\, d\alpha'$ to be zero, we should have from (10)

$$\text{(incomplete)} \quad \int^{n-2}\!\!\int \phi\, d\sigma \doteq \int^{n-1}\!\!\int \phi_9 \nabla_9 da, \quad . \quad . \quad . \quad (11)$$

exactly similar to (9). Unfortunately for simplicity, this is legitimate only in very special cases.

The true form of (11) may be obtained from our general theorem of integration, by which in particular

$$\int^{n-1}\!\!\int \dot{\psi}\, d\varsigma_2 = \int^{n-1}\!\!\int \psi_9 V_2 d\alpha \nabla_9. \quad . \quad . \quad . \quad (12)$$

Now, by (7) $d\sigma = V_1\iota d\varsigma_2$. Hence by (12)

$$\int^{n-2}\!\!\int \phi\, d\sigma = \int^{n-1}\!\!\int \phi_9 V_1 \iota_9 V_2 d\alpha \nabla_9. \quad . \quad . \quad (13)$$

It will be seen from this, since $da = V_0 \iota d\alpha$, that

$$\int^{n-2}\int \phi\, d\sigma - \int^{n-1}\int \phi_9 \nabla_9 da$$

instead of being zero, as stated in (11), consists of two terms,

$$-\int^{n-1}\int \phi_9 d\alpha V_0 \iota \nabla_9 - \int^{n-1}\int \phi V_1 (V_2 d\alpha \nabla \,.\, \iota),$$

the first involving the derivative $-V_0 \iota \nabla . \phi$ of ϕ along the tube ι, and the second the derivatives of ι in the form $-V_1(V_2 d\alpha \nabla . \iota)$. This last has a simple geometrical interpretation. Integrating it over a finite portion of $\int(d\alpha)$, we have by (12)

$$-\int^{n-1}\int V_1(V_2 d\alpha \nabla \,.\, \iota = -\int^{n-2}\int V_1 ds_2 \iota) = \int^{n-2}\int d\sigma$$

by (7). This geometrical result reminds us of the important fact that $\int^{n-2}\int d\sigma$. for a closed boundary, *in general differs from zero.* In a Euclidean space it would be zero.

Of special importance is the following particular case of (13) obtained by putting $\phi(\ ) = V_0 \bar{\tau}(\ )$, where τ is a contravariant vector density:

$$\int^{n-2}\int V_0 \bar{\tau} d\sigma = \int^{n-1}\int V_0 d\alpha V_1 \nabla V_2 \bar{\tau} \iota. \quad . \quad . \quad (14)$$

§ 3. *Proof of the two forms of Poisson's theorem.*—Let the position vector ρ of a point be given by

$$\left.\begin{aligned} \rho &= \sum_{c=1}^{n} x_c \iota_c, \\ \text{where} \qquad 1 &= -\iota_1^2 = -\iota_2^2 = \ldots = -\iota_{n-1}^2 = \mp \iota_n^2 \end{aligned}\right\}, \quad . \quad . \quad (15)$$

the upper or lower sign of ι_n^2 being taken according as the manifold is space-like or time-space-like.

To prove (1) take the point O as origin and use equiradial coordinates ["equiradial" seems a suitable term for the coordinates used in "A new Identity affecting Relativity," etc., Phil. Mag. †]. An asterisk attached to an equation signifies that it is true when the coordinates are equiradial, and in particular when they are Galilean. Thus

$$\left.\begin{aligned} h &= (-\rho^2)^{\frac{2-n}{2}}/(n-2), \\ \nabla h &= -(-\rho^2)^{-\frac{n}{2}}\rho, \\ \nabla^2 h &= 0, \end{aligned}\right\} [h \text{ given by } (1)]. \quad (16)^*$$

except for the point O itself. Draw a small sphere around the origin and consider its boundary as part of the integral boundary. With the origin thus excluded

$$\begin{aligned} 0 = \nabla^2 \int^n \int h\, db &= \int^n \int \nabla^2 h\, db, \\ &= \int^{n-1}\int V_0 d\alpha \nabla h. \end{aligned}$$

† [To appear shortly.]

[It is easy to see that $V_0 d\alpha \nabla h$ is the solid angle subtended by $d\alpha$ at O.] This result shows that the integral over an arbitrary boundary surrounding the origin is equal and opposite to that of the small sphere; or if both are taken in the same direction radially, they are equal. Now the value for the small sphere is π_n. The well-known argument of three dimensions holds, and we have that with the coordinates chosen

$$\nabla^2 \int^n \int h\overline{m}\, db = \pi_n \overline{m}. \quad . \quad . \quad . \quad . \quad (17)^*$$

This is the same as (1) in equiradial coordinates and the integral of (1) is an invariant. Hence (1) is true universally.

Similarly for (2). In the present case put $x_n = t$, and $\Sigma x\iota_1$ for $\sum_{c=1}^{n-1} x_c \iota_c$, and remember that

$$\nabla = \sum_{c=1}^{n-1} \iota_c \mathrm{D}_{x_c} - \iota_n \mathrm{D}_t. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Take O for the origin and use equiradial coordinates with axes so chosen that at the origin $\iota = \iota_n$. Thus

$$\left.\begin{array}{l} 2z = \Sigma x^2 - t^2, \ \nabla z = \rho, \ -V_0 \iota \nabla z = t, \\ h = (\Sigma x^2)^{\frac{3-n}{2}} / (n-3). \end{array}\right\} . \quad . \quad (19)^*$$

This is exactly the same form of h as we had in the proof of (1), except that here we have $n-1$, where before we had n. Hence

$$[h \text{ given by (2) and (5)}] \quad \nabla^2 h = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)^*$$

except where h is infinite, that is, on the axis of t. [As a matter of fact, our surface $\int (d\alpha)$, being space-like, cannot bend round and intersect the axis in a point other than O, so that O is the only singular point.]

We now have

$$\nabla^2 \int^{n-1} \int h\overline{m} V_0 \iota\, d\alpha = 4\pi \overline{m}. \quad . \quad . \quad . \quad (21)^*$$

because, first from (20), we may neglect all parts of the integral except near the origin, and there we may put $\iota = \iota_n$. The matter has now been reduced exactly to the former case, except that $x_1, x_2, \ldots x_{n-1}$ replace the former $x_1, x_2, \ldots x_{n-1}, x_n$. Hence (2) is true.

At this stage the following remarks seem desirable:—

(1) If the family and congruence have not been prescribed the family may be taken as the planes $x_n =$ const. in any given system of coordinates; that is, we may identify our present *coordinate* t with our former *parameter* t. At the same time we may take ι parallel everywhere to ι_n, which simplifies the differentiations when we transform (2), say by integration by parts. Instead, we might take $\eta\iota$ parallel

everywhere to ι_n. This would make the tubes everywhere orthogonal to the surfaces.

(2) The converse of (1) in all particulars holds. If both family and congruence have been prescribed, we can choose coordinates (flatten the surfaces and straighten the tubes), such that the family again becomes $x_n = \text{const.}$ and ι is everywhere parallel to ι_n: or instead, if only the family had been prescribed, we could have taken the tubes orthogonal to them, so making $\eta\iota$ parallel everywhere to ι_n.

(3) When the family and tubes are already prescribed, we can make $\iota = \iota_n$ everywhere, and one particular member of the family become $x_n = 0$. Of course, if that one member alone has been prescribed, we can now further take the whole family to be $x_n = \text{const.}$

We will conclude by proving Green's theorem. Let y and z be invariant functions of position. [When we require the original z of (2), (3), (4), we may use $Z = \frac{1}{2}$ (distance)^2.] In the space-like manifold integrate twice by parts, the following:

$$\int^n\!\!\int y \mathrm{V}_0 \nabla \eta \nabla z\, db = \int^{n-1}\!\!\int y \mathrm{V}_0\, d\alpha\, \eta \nabla z - \int^n\!\!\int \mathrm{V}_0 \nabla y \bar{\eta} \nabla z\, db$$
$$= \int^{n-1}\!\!\int \mathrm{V}_0\, d\alpha\, \eta (y \nabla z - z \nabla y) + \int^n\!\!\int z \mathrm{V}_0 \nabla \bar{\eta} \nabla y\, db.$$

Here we might have said instead: express

$$\int^{n-1}\!\!\int \mathrm{V}_0\, d\alpha\, \eta (y \nabla z - z \nabla y)$$

as a volume integral. Each of these processes can be imitated in the time-space-like manifold. Perhaps the second is easier to follow. Using (14) express

$$\int^{n-2}\!\!\int \mathrm{V}_0\, d\sigma\, \eta (y \nabla z - z \nabla y)$$

as an $\int^{(n-1)}\!\!\int (d\alpha)$. Thus we obtain the two forms of Green's theorem, *both* forms applying to *both* species of n-manifold,

$$\left.\begin{aligned} &\int^n\!\!\int (y \mathrm{V}_0 \nabla \bar{\eta} \nabla z - z \mathrm{V}_0 \nabla \bar{\eta} \nabla y)\, db \\ &\qquad = \int^{n-1}\!\!\int \mathrm{V}_0\, d\alpha\, \bar{\eta} (y \nabla z - z \nabla y), \\ &\int^{n-1}\!\!\int \mathrm{V}_0\, d\alpha (-y \mathrm{V}_1 \nabla \mathrm{V}_2 \iota \bar{\eta} \nabla z + z \mathrm{V}_1 \nabla \mathrm{V}_2 \iota \eta \nabla y \\ &\qquad\qquad - \bar{\eta} \mathrm{V}_1 \iota \mathrm{V}_2 \nabla y \nabla z) \\ &\qquad = \int^{n-2}\!\!\int \mathrm{V}_0\, d\sigma\, \bar{\eta} (y \nabla z - z \nabla y). \end{aligned}\right\} . \qquad (22)$$

The three dimensional form of (22) used to be known as Green's theorem. Jeans ('Electricity and Magnetism,' § 181) seems to say that Green's theorem, properly speaking, is what (9) and (14) become when we put

$$\phi = \mathrm{V}_0(\ )\bar{\tau},$$

where $\bar{\tau}$ is a contravariant vector density. In (22)

$$\tau - \eta(y \nabla z - z \nabla y).$$

University of Tasmania,
March 3rd, 1922.

XV. *The Mechanical Forces indicated by Relativity in an Electromagnetic Field. Can their existence be demonstrated?* By ALEX. McAULAY, *M.A., Professor of Mathematics, University of Tasmania* *.

§ 1. *EQUATIONS in C.G.S. measure.*—In the paper "Multenions and Differential Invariants," III. (P. R. S.) the electromagnetic field was treated on the lines of relativity confined to a Riemann manifold, on a new and, I hope, simplified basis. The special conclusions therefrom are, in the present paper, brought to a much more satisfactory halting-stage than in the former paper.

First it is desirable to indicate (see synopsis below) the changes of notation necessary to pass from rational to C.G.S. units.

t must everywhere be replaced by ct where c is the velocity of light *in vacuo*. This has to be remembered in every differentiation with regard to t; $\partial/\partial t$ must always be replaced by $c^{-1}\partial/\partial t$. There is one very important *apparent* exception to the statement that t is replaced by ct. Its cause may be thus stated. Merely to *preserve simplicity* in our mathematical expressions, $f\,db$ and $W\,db$ are no longer to be regarded as *elements of action*. They shall instead be taken to mean (action $\times c$). Thus

$$f \,.\, dx\, dy\, dz\, d(ct) = \text{action } \times c$$

and we derive the statement that

$$f \,.\, dx\, dy\, dz\, dt = \text{action}.$$

This is the apparent exception referred to, since it may seem here to be exceptionally imposed that t remains unchanged instead of being replaced by ct.

$\mathbf{R}$ is written for the quaternion $xi+yj+zk$, or, according to our identifications (see § 3 below) for the multenion $x\iota_2\iota_3+y\iota_3\iota_1+z\iota_1\iota_2$; $\acute{\upsilon}$ as usual is used for $\iota_1\iota_2\iota_3\iota_4$. Thus

$$\rho = \iota_4(\acute{\upsilon}\mathbf{R}+ct)$$

in C.G.S. units, instead of $\iota_4(\acute{\upsilon}\mathbf{R}+t)$ in rational units.

In C.G.S. units the $\mathbf{H}$ of rational units must be replaced by $(4\pi)^{-1}\mathbf{H}$, but we do not require to introduce a single other 4π.

The reader is recommended to think mainly of the E.M. system. Thus we shall write (c) and (c^{-1}) for what in a completely comprehensive scheme would be written as

$$(c) = c\mu = c^{-1}\mathrm{K}, \quad (c^{-1}) = (c)^{-1}.$$

* Communicated by the Author.

In my ordinary work I simply write c instead of (c) and c^{-1} instead of (c^{-1}), remembering that as a consequence, when I wish to pass to the E.S. system the only corresponding notations affected are the following :—

Rational system.	E.M. system.	E.S. system.
$\mathbf{E}$	$c^{-1}\mathbf{E}$	$c\mathbf{E}$
$\mathbf{D}$	$c\mathbf{D}$	$c^{-1}\mathbf{D}$

No changes whatever are required in the purely multenion symbols and equations, and none in the following quaternion symbols (vectors) :

$$\mathbf{R}, \mathbf{B}, \mathbf{K}, \mathbf{A}, \mathbf{F}.$$

Of the other quaternion symbols, $\mathbf{H}$ is replaced by $(4\pi)^{-1}\mathbf{H}$ as already remarked ; and

$$\mathbf{E}, \mathbf{D}, t, \mathrm{K}_4, \mathrm{A}_4, \mathrm{F}_4$$

are replaced respectively by

$$(c^{-1})\mathbf{E}, (c)\mathbf{D}, ct, c\mathrm{K}_4, c^{-1}\mathrm{A}_4, c^{-1}\mathrm{F}_4.$$

Synopsis of eqs. etc. in C.G.S. units.

Dimensional symbol.	Equations.	
L	$\rho = \iota_4(\acute{\upsilon}\mathbf{R} + ct)$	(1)
H	$\omega^{\bullet} = (4\pi)^{-1}\mathbf{H} + (c)\acute{\upsilon}\mathbf{D}$	(2)
B	$\omega^{\times} = \mathbf{B} + (c^{-1})\acute{\upsilon}\mathbf{E} = -{}_{\omega}\cdot\nabla f = \mathrm{V}_2\nabla\lambda^{\times}$ . .	(3)
X	$\kappa^{\bullet} = \iota_4(\acute{\upsilon}\mathbf{K} + c\mathrm{K}_4) = \mathrm{V}_1\nabla\omega^{\bullet}$	(4)
P	$\lambda^{\times} = \iota_4(\acute{\upsilon}\mathbf{A} + c^{-1}\mathrm{A}_4)$	(5)
F	$\nu^{\bullet\times} = \iota_4(\acute{\upsilon}\mathbf{F} + c^{-1}\mathrm{F}_4) = \mathrm{V}_1\kappa^{\bullet}\omega^{\times}$	(6)

Notes on the synopsis.—Dimensional symbol F is not force but force per unit volume. For consistency of notation $\lambda^{\times}$ (not used in our work) has been written for our previous λ. In *(6)*, on our theory, $\nu^{\bullet\times}$ does not represent the whole mechanical force ; see expressions (9), (9 *a*), (9 *b*) of § 2 below. The (4π)'s are seen to present no trouble *when* permittivity K and permeability μ are absent. [In our scheme above we should replace K, μ of rational system by $(4\pi)^{-1}$K, $4\pi\mu$ (to pass to C.G.S. units) respectively.]

It is well to add a seventh equation. ∇ in the multenion sense and ∇ in the quaternion sense are mathematically different. Putting ∇' for the quaternion ∇ we have in C.G.S. units

$$\nabla = \iota_4(\acute{\upsilon}\nabla' - c^{-1}\mathrm{D}_t). \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In rational units, of course, D_t is written in place of the $c^{-1}D_t$ of (7).

∇ is of dimensions L^{-1} and ${}_{\omega}\cdot\nabla$ of dimensions H^{-1}. Hence from the synopsis eqs. the dimensional relations are seen to be

$$HB = LF = (f), \quad X = L^{-1}H, \quad P = LB, \quad . \quad . \quad (8)$$

where (f) is the dimensional symbol of f. The reader will easily verify from these statements that all our equations are correct dimensionally. *Note.*—HB is in ordinary theory energy per unit of three-dimensional volume. Now for simplicity we desire $\omega^{\bullet}$, $\omega^{\times}$ to be of the same dimensions as **H**, **B** respectively; and *also* we desire to retain the simple equation

$$\omega^{\times} = -{}_{\omega}\cdot\nabla f.$$

This requires $f\,dx\,dy\,dz\,d(ct) = f\,db$ to be of dimensions $c \times$ action as we prescribed above. It will be seen that our prescriptions have left the dimensions of H *or* B absolutely arbitrary; with regard to H and B we have prescribed *only* that HB is force per unit area, that is energy per unit volume (area and volume in the usual three-dimensional sense).

§ 2. *The stresses and forces in an electromagnetic field.*—In § 21 of "Mult. etc." III., in the equations (23 *a*), (19 *a*), and in the equations (24), (25), (26), when supplemented by the expressions (24 *a*), (25 *a*), (26 *a*), a vector, normal to ι, occurs and is denoted by $\xi\cdot\iota = {}_{\iota}\nabla f$. We shall here denote it by $\sigma^{\bullet\times}$. The vector can be completely evaluated in terms of $\omega^{\bullet}$, $\omega^{\times}$.

From eq. (26), when supplemented, we have

$$2_{\theta}\mathrm{Cl}\,f\alpha^{\times} = \theta^{-1}\alpha^{\times}\,.\,f - \theta^{-1}V_1(V_3\omega^{\times}\alpha^{\times})\omega^{\bullet} - \theta^{-1}\sigma^{\bullet\times}\,.\,V_0\iota\alpha^{\times}.$$

$2_{\theta}\mathrm{Cl}\,f - f\,.\,\theta^{-1} = \phi$ is self-conjugate, so that $V_2\zeta\phi\zeta = 0$. Using this we obtain at once

$$V_2\iota\vartheta^{-1}\sigma^{\bullet\times} = V_2(V_1\zeta\omega^{\bullet})\theta^{-1}(V_1\zeta\omega^{\times}). \quad . \quad . \quad . \quad (1)$$

Since $\sigma^{\bullet\times}$ is normal to ι

$$\sigma^{\bullet\times} = V_1\iota\theta V_2\iota\theta^{-1}\sigma^{\bullet\times}$$

or

$$\sigma^{\bullet\times} = V_1\iota V_2(\theta V_1\zeta\omega^{\bullet})(V_1\zeta\omega^{\times}). \quad . \quad . \quad . \quad . \quad (2)$$

This is the really useful form of $\sigma^{\bullet\times}$ since it gives in neutral form

$$\text{(neutral form)} \quad \sigma^{\bullet\times} = V_1\iota V_2\omega^{\bullet}\omega^{\times}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

but if we please we can rid eq. (2) of ζ, obtaining

$$\left.\begin{aligned}\sigma^{\bullet\times} &= \theta V_1(V_1\omega^\times\iota)\omega^\bullet - V_1(V_1\omega^\bullet\theta\iota)\omega^\times \\ &= \theta V_1\omega^\times V_3\iota\omega^\bullet \quad - V_1\omega^\bullet V_3\theta\iota\omega^\times\end{aligned}\right\} . \quad . \quad . \quad (4)$$

In Galilean coordinates we have from (3)

$$\sigma^{\bullet\times} = V_1[V_2(\mathbf{BH}+\mathbf{DE}) - \acute{\upsilon}V_2(\mathbf{DB}-\mathbf{EH})]\iota. \quad . \quad . \quad (5)$$

[In § 3 below is explained in detail how readily to effect the reductions of this § 2. We have thought this procedure would render the main argument clearer.] Thus putting

$$\iota = \iota_4(\acute{\upsilon}\mathbf{V}+1)/\sqrt{(1+\mathbf{V}^2)}, \quad . \quad . \quad . \quad . \quad (6)$$

so that $\mathbf{V}$ is the velocity, we obtain

$$\sigma^{\bullet\times}\sqrt{(1+\mathbf{V}^2)} = \iota_4\{\acute{\upsilon}(V[V(\mathbf{BH}+\mathbf{DE}).\mathbf{V}] \\ + V[\mathbf{DB}-\mathbf{EH}]) - S(\mathbf{DB}-\mathbf{EH})\mathbf{V}\}. \quad . \quad . \quad (7)$$

Eq. (19 a) reads

$$-lT_{(\zeta)}\zeta = V_1\kappa^\bullet\omega^\times - \nabla_9 V_0\iota_9\sigma^{\bullet\times} - \sigma_9{}^{\bullet\times}V_0\iota_9\nabla_9. \quad . \quad (8)$$

Our most important deduction from this happens to be easy; so before treating (8) generally we will make this deduction. Let the coordinates be chosen so that $\mathbf{V}$ is zero at the point considered and therefore ι becomes ι_4. In the neighbourhood of the point let the medium also be at rest; in other words let the point be in a rigid body (or, at any rate, a body wholly at rest). Thus ι is the constant ι_4 and all its derivatives are zero; and $-V_0\iota\nabla = D_t$ and $\mathbf{V}$ of (7) is zero. Thus the quaternion expression for the force per unit volume is

$$V\mathbf{KB} - \mathbf{E}S\nabla\mathbf{D} + D_t V(\mathbf{DB}-\mathbf{EH}). \quad . \quad . \quad . \quad (9)$$

The third term expresses a hitherto unsuspected (I believe) mechanical force exerted by the field. Can any physicist verify its existence? In C.G.S. measure the term is

$$D_t V(\mathbf{DB} - \mathbf{EH}/4\pi c^2).$$

[This is a case of the remark that to pass to C.G.S. units D_t must be replaced by $c^{-1}D_t$.]

Returning to (8), since $V_0\iota\sigma^{\bullet\times} = 0$,

$$-\nabla_9 V_0\iota_9\sigma^{\bullet\times} = \nabla_9 V_0\iota\sigma_9{}^{\bullet\times}.$$

Hence putting $\sigma^{\bullet\times} = l\sigma^\times$ and $l\iota = \iota^\bullet$ (8) may be easily written in the invariantive form

$$-lT_{(\zeta)}\zeta = V_1\kappa^\bullet\omega^\times + V_1(V_2\nabla\sigma^\times)\iota^\bullet - \sigma^\times V_0\nabla\iota^\bullet. \quad . \quad (10)$$

Each of the two terms in $\sigma^{\times}$ is separately normal to ι. That this portion of the force is normal to ι means that the corresponding portion of $F_4(=-S\mathbf{EK})$ is zero when the medium is taken to be at rest; that is, there is no unsuspected energy transformation of an electrical or heat generating nature, due to the new term in the mechanical force.

In interpreting (10) assume, at the point considered, $\mathbf{V}$ to be zero, but neither $D_t\mathbf{V}$ nor $(\nabla_9, \mathbf{V}_9)$ to be zero. The middle term of (10) thus contributes the third term of (9) together with

$$\left.\begin{array}{c} V[V(\mathbf{BH}+\mathbf{DE})]D_t\mathbf{V}'-\nabla_9 S(\mathbf{DB}-\mathbf{EH})\mathbf{V}_9' \\ \mathbf{V}'=\mathbf{V}/\sqrt{(1+\mathbf{V}^2)} \end{array}\right\} \quad . \quad . \quad (9\,a)$$

where

The third term of (10) contributes

$$V(\mathbf{DB}-\mathbf{EH}).(-S\nabla\mathbf{V}'+D_t[1/\sqrt{(1+\mathbf{V}^2)}]). \; . \quad (9\,b)$$

In changing to C.G.S. units $\mathbf{V}$ must, of course, be replaced by $c^{-1}\mathbf{V}$.

I imagine these terms depending on acceleration and time-rate of strain are much too small to be determined experimentally. The physicist who understands the meaning of three-dimensional vectors has been placed by the expressions above in a position to judge for himself.

In addition to the three terms (9), (9 a), (9 b), expressing the mechanical force when the medium is at rest there will, of course, be terms due to the coordinates of strain entering the expression for action. What has been established is that a *perfectly arbitrary* invariant function $f(\omega^{\bullet}, \omega^{\bullet}, \theta)$ occurring as the field terms in W (action per unit of four-dimensional volume), a *quite definite* force per unit volume (from which the arbitrary function has completely vanished) must exist. The function, of course, has its effects, but these are merely to fix the details of interrelation of **B**, **H**, **D**, **E**. This result seems very surprising. Perhaps relativity will give similar guidance when elasticity etc. has been brought under its treatment.

§ 3. *Transformation from multenion to quaternion forms.*—With our identifications, a real quaternion Q is taken to mean

$$Q=a\iota_2\iota_3+b\iota_3\iota_1+c\iota_1\iota_2+g$$
$$=ai+bj+ck+g=\mathbf{G}+g,$$

and an imaginary quaternion $\acute{\upsilon}Q'$ to mean ($\acute{\upsilon}=\iota_1\iota_2\iota_3\iota_4$)

$$\acute{\upsilon}Q'=a'\iota_4\iota_1+b'\iota_4\iota_2+c'\iota_4\iota_3+g'\iota_1\iota_2\iota_3\iota_4$$
$$=\acute{\upsilon}(a'\iota_2\iota_3+b'\iota_3\iota_1+c'\iota_1\iota_2+g')=\acute{\upsilon}(\mathfrak{G}'+g'),$$

$Q+\acute{\upsilon}Q'$ is a bi-quaternion (in Hamilton's sense). The general multenion q is expressed by means of two bi-quaternions $Q+\acute{\upsilon}Q'$, $P+\acute{\upsilon}P'$ and ι_4 thus

$$q=(Q+\acute{\upsilon}Q')+\iota_4(P+\acute{\upsilon}P').$$

Let a be an arbitrary real scalar and **A**, **B** arbitrary real vectors (of the quaternion system). Then we have the following scheme :—

Forms of V_cq ; $c=0, 1, 2, 3, 4$.

$$V_0q=a \qquad V_1q=\iota_4(\acute{\upsilon}\mathbf{A}+a) \qquad V_2q=\mathbf{A}+\acute{\upsilon}\mathbf{B}$$
$$V_4q=\acute{\upsilon}a \qquad V_3q=\iota_4(\mathbf{A}+\acute{\upsilon}a)$$

It is easy to pick out the five parts of q given by the last equation. Thus

$$V_0q=SQ, \qquad V_1q=\iota_4(\acute{\upsilon}VP'+SP), \qquad V_2q=V(Q+\acute{\upsilon}Q'),$$
$$V_4q=\acute{\upsilon}SQ', \qquad V_3q=\iota_4(VP+\acute{\upsilon}SP').$$

The one biquaternion furnishes the even parts and the other (the ι_4 biquaternion) the odd parts, thus

$$Q+\acute{\upsilon}Q'=(V_0+V_2+V_4)q,$$
$$\iota_4(P+\acute{\upsilon}P')=(V_1+V_3)q.$$

The imaginary $\acute{\upsilon}$ and also ι_4 are both commutative with the quaternions, but they are anti-commutative with one another, or

$$\iota_4\acute{\upsilon}=-\acute{\upsilon}\iota_4.$$

As a general rule collect all the ι_4's at the left of a complicated quadri-quaternion expression, attending to change of sign at each passage of ι_4 past $\acute{\upsilon}$. Then use the scheme for further treatment.

These remarks receive illustration in the transformations of § 2.

$$V_1\kappa\cdot\omega^{\times}=V_1\iota_4(\acute{\upsilon}\mathbf{K}+K_4)(\mathbf{B}+\acute{\upsilon}\mathbf{E}).$$

Consulting the scheme under V_1 we have from the bi-quaternion $(\acute{\upsilon}\mathbf{K}+K_4)(\mathbf{B}+\acute{\upsilon}\mathbf{E})$ to pick out the part of the

form $\acute{\upsilon}A+a$, that is,

$$\acute{\upsilon}(\mathrm{V}\mathbf{KB}+\mathrm{K}_4\mathbf{E})-\mathrm{S}\mathbf{KE}.$$

The rest of the bi-quaternion merely gives $\mathrm{V}_3\kappa^{\bullet}\omega^{\times}$ with which we have no concern. Thus

$$\mathrm{V}_1\kappa^{\bullet}\omega^{\times}=\iota_4(\acute{\upsilon}[\mathrm{V}\mathbf{KB}+\mathrm{K}_4\mathbf{E}]-\mathrm{S}\mathbf{KE}).$$

Hence we have the first two terms $\mathrm{V}\mathbf{KB}+\mathrm{K}_4\mathbf{E}$ of (9).

Next deduce (5) from (3)

$$\begin{aligned}\sigma^{\bullet\times}&=\mathrm{V}_1(\mathrm{V}_2\omega^{\times}\omega^{\bullet})\iota,\\&=\mathrm{V}_1[\mathrm{V}_2(\mathbf{B}+\acute{\upsilon}\mathbf{E})(\mathbf{H}+\acute{\upsilon}\mathbf{D})]\iota,\\&=\mathrm{V}_1[\mathrm{V}_2(\mathbf{BH}+\mathbf{DE})-\acute{\upsilon}\mathrm{V}_2(\mathbf{DB}-\mathbf{EH})]\iota.\end{aligned}$$

Next deduce (7) using the abbreviations

$$\mathbf{V}'=\mathbf{V}/\sqrt{(1+\mathbf{V}^2)},\quad \mathbf{X}=\mathrm{V}(\mathbf{BH}+\mathbf{DE}),\quad \mathbf{Y}=\mathrm{V}(\mathbf{DB}-\mathbf{EH}),$$
$$1/\sqrt{(1+\mathbf{V}^2)}=b,$$

so that b is what is usually denoted by β. We have

$$\begin{aligned}\iota&=\iota_4(\acute{\upsilon}\mathbf{V}'+b),\\\sigma^{\bullet\times}&=\mathrm{V}_1(\mathbf{X}-\acute{\upsilon}\mathbf{Y})\iota_4(\acute{\upsilon}\mathbf{V}'+b),\\&=\mathrm{V}_1\iota_4(\mathbf{X}+\acute{\upsilon}\mathbf{Y})(\acute{\upsilon}\mathbf{V}'+b).\end{aligned}$$

From the biquaternion ( ) ( ) following ι_1 we have to pick out the part of form $\acute{\upsilon}\mathbf{A}+a$. Thus

$$\sigma^{\bullet\times}=\iota_1[\acute{\upsilon}(\mathrm{V}\mathbf{XV}'+b\mathbf{Y})-\mathrm{S}\mathbf{YV}']$$

which is eq. (7).

Similarly (11) and (12) can be deduced from (10).

In conclusion it may be remarked that

$$a\iota_1+b\acute{\upsilon}+c\iota_1\acute{\upsilon}+g$$

is itself a quaternion such that we may put

$$\iota_4=i\sqrt{(-1)},\quad \acute{\upsilon}=j,\quad \iota_4\acute{\upsilon}=k\sqrt{(-1)}.$$

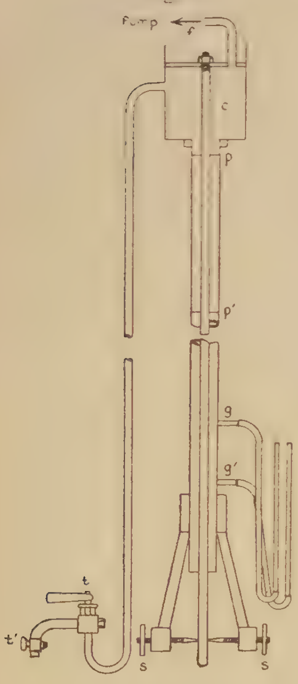
University of Tasmania,
28 November, 1922.

XVI. *The Flow of Water in the Annular Space between two Coaxal Cylindrical Pipes.* *By* THOMAS LONSDALE, *B.Sc.**

SEVERAL investigations have previously been made on the flow of liquids and gases in cylindrical pipes. So far as I know, no similar investigations have been made on fluids flowing in the annular space between two coaxal pipes. The present paper deals with an investigation for this special case. Preliminary experiments showed that it was necessary to have the pipes in a vertical position instead of the more usual horizontal one, owing to the impossibility of counteracting the tendency of the inner pipe to sag inside the outer pipe when the pipes were in the latter position. The general nature of the flow was much the same as in the cylindrical pipe, there being a well-defined critical velocity separating regions of turbulent and non-turbulent flow.

Apparatus and Method.

Fig. 1.



The form of apparatus used is shown in fig. 1. The water was obtained from a tank in the roof of the laboratory, and the flow was controlled by the tap *t*, small variations being obtained by the means of side tap *t′*. From the tap the water passed into the canister *c* and then down the annular space. A filter pump *f* removed the air from *c* when starting up, and was thereafter kept running. Pegs *p* and *p′* attached to the inner cylinder served to centre it at the top and about a third of the way down, and the adjustable screws *s*, pointed so as to disturb the flow as little as possible and acting upon the protruded end of the inner tube, centred it at the lower end of the outer tube. The velocities of flow of the water were obtained by letting it run into a calibrated tank for a measured time.

* Communicated by Prof. A. W. Porter, D.Sc., F.R.S.

The pressures at two points in the tube from which the pressure gradient was obtained were measured with the water manometers qq'. It is remarked by Stanton and Pannell (Phil. Trans. Roy. Soc. A. 214) that in the case of flow in a pipe the length of "leading in" pipe through which the water must pass to enable irregularities of flow due to the inlet to be damped out varies from 90 to 140 d where d is the diameter. In the present experiments the distance between the pegs p' and the first gauge hole q was about $100\,(b-a)$ where b and a are the diameters of the outer and inner cylinders. The distances between q and q' and between q' and the orifice of the outer tube were about $10(b-a)$. The diameters of the cylinders experimented on are given in Table I. The two larger outer tubes were of smooth drawn brass and the smaller of smooth drawn copper. The 1·90 and 1·29 inner cylinders were tubes of a nickel-silver alloy mounted on mild steel shafting, and the ·79 cylinder was a drawn phosphor-bronze rod.

No one material was found to have the necessary straightness in all the lengths and diameters required, and it was thought that the closer approximation to true annularity that was obtained by the use of varying materials, all smooth-drawn by the same process, would more than counterbalance the error due to possible differences created in the surface friction.

The gauge holes in the outer cylinder were made in the usual manner. The temperature of the water was observed by means of a thermometer placed in the water flowing to waste. In the preliminary experiments no effective difference was found between this temperature and that of the water in the canister. This temperature did not vary by more than ·5° C. throughout a set of observations. On each system, sets of observations on the relation between the pressure gradient and the mean velocity were taken in the directions of increasing flow and decreasing flow. During the preliminary work it was found that for low velocities tiny bubbles of air collected on the walls of the annular space. The rate of accumulation was small, but was enough to produce an effect equivalent to an increase in diameter of the inner cylinder. This seemed to be due to the siphon-like form of the apparatus. In order to obviate this difficulty, after each reading had been taken the tap was turned full on in order to drive all the air out, and was then adjusted until the gauges stood at their former levels. The flow was then increased or decreased according to the direction of procedure, and

another reading was taken. A slight modification was made in going in the direction of increasing flow from the non-turbulent to the turbulent region. In the neighbourhood of the critical velocity, after a reading had been taken and the the air had been driven out, the flow was decreased until the gauges indicated a velocity well in the non-turbulent region. The flow was then very carefully increased until the previous gauge levels had just been passed, and the flow corresponding to the new level was measured. This was done in order to detect any persistence of stream-line flow which would give a critical velocity analogous to the second critical velocity of Osborne Reynolds (Phil. Trans. Roy. Soc. 1883). No such effect was observed. In the region just above the critical velocity where the gauges were very unsteady, the amplitude of movement was observed and the mean taken as indicating the pressure.

Discussion of Results.

From the measurements taken we can obtain the law of variation of the pressure gradient with the velocity in the turbulent region, and a formula determining at what velocity turbulence will begin. We will consider these two parts in succession.

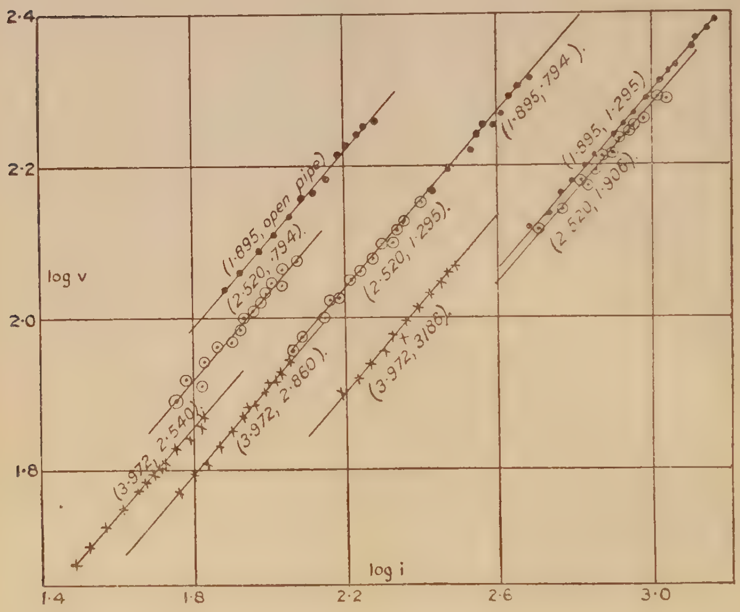
Turbulent Region.—On plotting the logarithms of pressure gradient and mean velocity in the turbulent region for each system, curves were obtained as shown in fig. 2. The figures in brackets beside each curve indicate the diameters of the outer and inner cylinders of the system for which the curve is a representation of the relation between the log pressure gradient and the log velocity.

Examination reveals the facts that the curves on the diagram are linear, and that the slope of the lines indicates that if i is the pressure gradient and v the velocity, then $i=Cv^n$, where $n=1{\cdot}72$ and C is a constant for any particular system. Furthermore, from fig. 2 we see that the curves are nearly superposed for cases where $b-a$ is nearly the same (where b and a are the diameters of the outer and inner cylinders respectively). This suggests that C contains a function of $b-a$ as a factor: we will assume that a simple power function will suffice. An application of the method of dimensions leads to the expression

$$i = \mathrm{K}(b-a)^{n-3}\mu^{2-n}\rho^{n-1}v^n,$$

where μ is the viscosity and ρ the density of the liquid. This is of the usual form for an ordinary pipe where the range is sufficiently restricted for an index law to hold for the velocity, the only difference being that $(b-a)$ replaces

Fig. 2.



the diameter of the pipe. In Table I. are shown the values of K calculated from my observations for the case $n=1\cdot72$. Readings were also taken upon the 1·79 cm. ordinary pipe, in order to compare them with the results of other experimenters. The results in square brackets are computed for comparison from Stanton and Pannell's observations for smooth brass pipes (Phil. Trans. Roy. Soc. A. 214).

It will be seen that K for the systems where $b=2\cdot52$ cm. is larger than in the other cases. This may be due to the 2·52 cm. pipe being slightly rougher than the other two outer pipes. It seems evident that the formula put forward gives a fair approximation to the experimental results. The mean value for K is ·214. The units used are the C.G.S. system. The pressure gradient i is measured in dynes

TABLE I.

The Space between two Coaxal Cylindrical Pipes.

b.	a.	$b-a$.	K.
1·895	1·295	0·60	·212
1·895	0·794	1·101	·206
1·895	0·019	1·876	·199
2·520	1·906	0·614	·228
2·520	1·295	1·225	·231
2·520	0·794	1·726	·225
3·972	3·186	0·786	·211
3·972	2·860	1·112	·197
3·972	2·540	1·432	·217

Ordinary Pipe.

b.	K.
1·895	·197
[1·255]	[·210]
[2·855]	[·213]

per cm.^{3} and the velocity v in cm. per sec. The limits over which the formula has been tested experimentally should be made quite clear. The maximum velocities obtained were about 2·5 metres per second, and the largest pipe used had a diameter of 3·97 cm. The largest difference in diameters was 1·73 cm.

Critical Velocity.—In the non-turbulent region the relation between the velocity of flow and the pressure gradient is given by

$$v = \frac{1}{32\mu} i \left\{ b^2 + a^2 - \frac{b^2 - a^2}{\log \frac{b}{a}} \right\}^{*}.$$

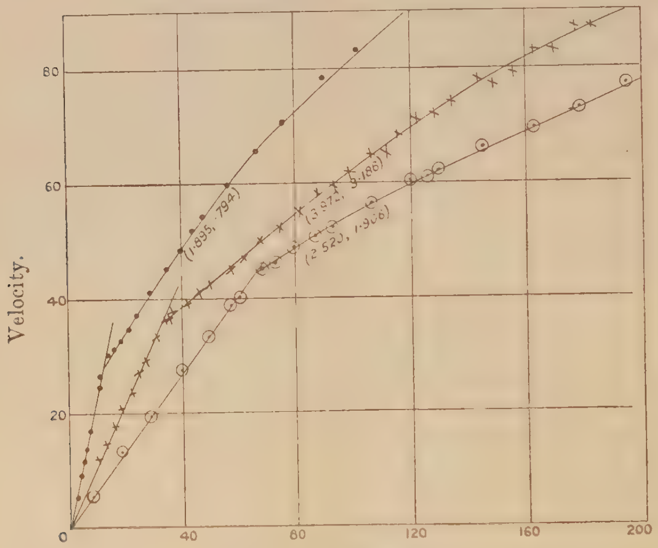
In the non-turbulent region this relation was found to be linear, as indicated by the formula, but when the critical velocity was just exceeded, the increase in i became large for a small increase in v, and the i, v curve bent sharply away from the straight line. Fig. 3 shows the type of curves obtained. The critical velocity is given quite definitely. In the case of the ordinary pipe this velocity v_c is given by $v_c = \frac{K\eta}{d}$, where η is the "kinematical viscosity" or the viscosity of the fluid divided by its density; d is the diameter and K is a constant. This constant varies according to the nature of the experiments and the materials of which the pipe is composed. In the papers previously referred to, Osborne Reynolds finds K$=$2000 for drawn lead

* Lamb, 'Hydrodynamics,' 4th edit. p. 579.

pipes, while Stanton and Pannell find K=2500 for drawn brass pipes.

The principle of Dynamical Similarity (Stanton and Pannell, *loc. cit.*) can give us no information in the case of the flow of liquids in the annular space between two coaxal cylinders, because the linear dimensions of each system cannot be expressed by a single parameter corresponding

Fig. 3.



to the diameter in the case of the ordinary pipe. It is thus necessary to fall back on some purely empirical relationship to express the variation of the critical velocity with the diameters of cylinders in each system. For any one system v_c is found to vary directly with η, as in the case of the ordinary pipe. Experiments made on two systems in summer and winter showed that for a 30 per cent. change in η the ratio changed by less than 2 per cent.

For observations at the same temperature on different annular systems, it appears that at the critical velocity for any system the *quantity* of liquid is, roughly, directly proportional to the outer diameter of the system and independent of the inner diameter.

Expressing this in symbols, we have $v_c = K\eta \frac{b}{b^2 - a^2}$, where K is a constant having the value just under 4000, and $b^2 - a^2$ is proportional to the cross-sectional area.

The variation of the experimental results from any systematic empirical relationship does not justify the putting forward of a more elaborate expression. Table II. shows the experimental values of $\frac{v_c}{\eta}$ for each system and the values given by the formula putting K = 4000. Each experimental value represents the mean of eight determinations, which differed amongst themselves by about 5 per cent.

TABLE II.

b in cm.	a.	$\frac{v_c}{\eta}$ (experiment).	$\frac{v_c}{\eta}$ (formula).
1·895	1·295	3230	3970
1·895	0·794	2650	2560
2·520	1·906	3970	3700
2·520	1·295	2210	2160
2·520	0·794	1620	1760
3·972	3·186	2650	2820
3·972	2·860	1780	2090
3·972	2·540	1840	1700

The only serious divergence between formula and experiment, bearing in mind the variability of critical velocity measurements, occurs in the first line. This may be supposed to be due to a lack of axial symmetry in this system. Owing to the small diameter of the cylinders and the smallness of the annular space a divergence from axial symmetry is more likely to occur in this case than in any other, small diameter tubing being proportionally less straight than that of large diameter. The cases of the 1·89 open pipe and the (1·89, ·019) system could not be investigated because the trumpet-shaped orifice stream became unstable and broke up before the critical velocity was reached.

In conclusion, I should like to express my thanks to Prof. A. W. Porter, F.R.S., for the many valuable suggestions that he has made and the encouragement that he has given me during the course of the work, and to my father, Mr. J. J. Lonsdale, D.Sc., for bearing the cost of the investigations.

XVII. *Movements of the Earth's Surface Crust, II.* By J. JOLY, *F.R.S.**

IN a recent paper appearing in this Journal (June, 1923) I dealt with a theory of the source of the surface changes experienced by the Earth over Geological time. In this present paper I propose to supplement the references to (*a*) the conditions of equilibrium leading to transgressional seas, and (*b*) the distribution of temperature in the continental crust.

Transgressional Seas.

In my former paper (*loc. cit.*) the development during inter-revolutionary times of a sub-oceanic crust was referred to. The conclusion was reached that such a crust might attain a thickness of some 15 miles (24 kilometres) in 25 million years.

Now the development of such a crust must have the effect of securing to the ocean a basal support extending nearly as far downwards into the substratum as the average basal level of the continental layer.

In order to arrive at this level we have to decide upon the most probable depth of the continental layer. Seismology has led to estimates between 30 and 35 kilometres. I shall take 32 k. If, now, we add the mean continental elevation (700 m.) to the mean depth of the ocean (4400 m.) and subtract from 32 k., we find the submergence of the continents in the substratum to be 27 k. The surface of the ocean floor is taken as the upper surface of the substratum. The base of the sub-oceanic crust is 24 kilometres below this level.

Now when fusion of the magma becomes widespread and its voluminal expansion lifts both ocean and continents together, there can be no general differential vertical movements due to loss of buoyancy, save in so far as the continental base may project into the magma below the general basal level of the sub-oceanic crust. This, on our present reckoning, will be only some 3 kilometres. It may be nil. But on the other hand, the deeply projecting

* Communicated by the Author.

compensations will experience the effects due to the lessening density of the magma. They will grow heavier, as it were; the former isostatic equilibrium will be disturbed; and, locally, the continents will sink or sag downwards.

This effect will mainly take place where great mountain ranges and raised plateaux exist at the surface. Such transgressional flooding as affected the North American Continent during Laramide times would owe its initiatory development to this source. Similarly the earlier enlargement of the Mediterranean would be associated with the older mountain ranges then existing. In short, it is to such effects that we must ultimately ascribe the phenomenon of mountains begetting mountains. For such hollows must collect the débris of the millions of years which follow.

As the period of revolution approaches, the ocean floor begins to melt away and commingle with the general magma. Increasing circulation, both lateral and vertical, assails it with hot and, possibly, superheated currents. As it dwindles the continents become more and more exposed to the molten lava, and, as they had attained their former elevation relatively to the ocean at a period when the lava was at its maximum density, they must now experience as a whole the effects of the diminished buoyancy and begin to sink relatively to the ocean level. The ocean cannot experience this effect, but its floor may buckle or subside. To such movements the diminished rigidity of the suboceanic crust would contribute and, subsequently, the lateral compression to which it would be subjected; as referred to in my first paper. The deeps near continental margins are probable testimony to such compressional stresses. We possess unassailable evidence of vertical movements of the ocean floor even within recent times.

It is evident that with such a complexity of factors and without any sure knowledge of the amount and downward extent of the density change of the substratum, estimates of the differential vertical movements finally reached cannot be of value. It seems certain, however, as stated in my former paper, that the movements must be adequate to account for such estimates of the depths of transgressional seas as Geologists have been able to arrive at.

We possess, indeed, an indication of the prevailing relative densities of the submerged continental materials and the sustaining magma which is of special interest.

If it be accepted that the continental emergence is about 5000 m. and its submergence 27,000 m. then the ratio of

the density of the average submerged continental materials to that of the magma must be as 27 : 32 or as 2·53 : 3·0. A correction for the buoyant effects of the ocean will bring the ratio to about 2·65 : 3·0; which is in satisfactory agreement with the fundamental assumption that the continents do in fact, float on a substratum of basaltic magma.

In order to illustrate what has been said, let us consider the case of the continent of Africa. The mean height of that continent is stated to be 732 m. over sea level. The great plateau extending over its southern, central and eastern regions, possesses a mean altitude of 1332 metres. It, therefore, rises 600 m. over the general continental surface. If the density ratio of continental crust and magma is 2·6 to 3·0 we find that a compensating protuberance must extend nearly 4 kilometres into the magma to fulfil the condition that equal mass must underlie equal areas.

Now if the density of the substratum changes 10 per cent. a downward movement of the plateau region of 400 m. must occur to restore isostatic equilibrium. The force so originating will be supplemented by similar isostatic forces due to such compensations as may exist beneath the Atlas ranges and the Abyssinian region. The effects of the vertical stresses may be to depress the continent—possibly tilting it—so that transgressional waters will invade its low-lying regions. Thus the desert regions to the north-east, now at an altitude of about 300 metres over sea-level, would be flooded. Again they might possibly give rise to rifting of the continent in such direction as would most relieve the stresses. It seems probable that this is the sort of effect which must usher in a revolution.

One outcome of the foregoing views is the recognition of the fact, that cyclical changes of stress must, from the earliest times, have affected the earth's surface crust. During the period of advancing liquefaction the outer crust, as a whole, must experience stresses of tensile character. For the earth's surface is then increasing in area. When the climax of revolution is attained there must be a period of relaxation and recovery. Following this the shrinking of the substratum inaugurates a period of compressive stresses. This is the orogenic period as explained in my last paper. These stresses also die out and a long period of comparative repose attends the slow accumulation of radioactive heat in the substratum.

These cyclical changes of stress must have profoundly modified the surface features of the earth. They are the inevitable outcome of the existence of the substratum and the presence of radioactive elements throughout the materials of the earth's surface.

The Distribution of Temperature in the Continental Crust.

It follows from the principles of isostasy that the thickness of the continental crust must vary considerably; extending downwards inversely as the greater surface features extend upwards. Taking the average density of the submerged continental materials as 2·6 and that of the substratum as 3·0 (*ante*), the compensations must extend downwards 6·5 times the elevation above the mean level of the raised surface features.

For the radioactivity of the continental crust we have a certain choice of data according as we take it to be of acid or of intermediate character. I shall assume that it possesses a radioactivity as if it were compounded of equal amounts of acid and intermediate rocks. This will be found to involve the development of $0{\cdot}27 \times 10^{-12}$ calorie per gram per second *. If the density be 2·6 this becomes $0{\cdot}70 \times 10^{-12}$ cal. per sec. per c.c.

I shall first consider the question of the distribution of temperature in the average continental layer. In my last paper (*loc. cit.* p. 1174) I gave a computation showing that for a crust of 24 kilometres thickness, the steady output of radioactive heat must equal that indicated by the surface gradient. This is, however, defective not only in under-estimating the thickness of the continental layer but in under-estimating the heat flow indicated by the assumed surface gradient. It is, therefore, necessary to consider the question afresh.

If we adhere to our former estimate of continental thickness, *i. e.*, 32 k. and take the average conductivity as 4×10^{-3} the basal temperature is found to be closely 900° C. (Strutt, Proc. R. S. 77 A). This, in the first place, is a sufficient approximation to the temperature of the substratum to meet the real point at issue:—that escape of heat through the continental stratum cannot take place to any considerable

* Phil. Mag. Oct. 1912 & April 1915, also June 1923.

extent. With a temperature difference of only 250° the leakage would be small and would be closed by the growth of a thin basaltic crust congealed on the continental base. In point of fact had we assumed a continental thickness of 35 k. (which would have sufficiently satisfied the views of seismologists) and 'acid' radioactivity (also permissible), the calculated basal temperature would have come out as 1225°. The value selected above for the conductivity is the mean for granites, 'whinstones,' micaschists, and 'traps' as cited by Everett (C.G.S. System of Units).

We have now to consider how far this result may agree with gradients as observed in bore-holes, etc. over the continental surface. The heat coming to the surface originates in two ways. One part is generated by the radioactivity of the continental layer; the other ascends from beneath. The first is readily calculable (on data already cited) as 224×10^{-8} cal. per sec. per sq. cm. The second is 31×10^{-8} cal. per sec. per sq. cm. The total is $2{\cdot}55 \times 10^{-6}$ cal.

Thermal gradients, as all know, are very various—ranging from 26 to over 54 metres per degree centigrade. They steepen with depth so that in a distance downwards of 1000 metres they may steepen from 49 m. to 29 m.* This fact, which is of general occurrence, shows that either some loss of sensible heat takes place or that, approaching the surface, the conductivity increases—as, for instance, due to presence of water.

This gain in conductivity due to moisture is exhibited in experimental determinations. We find dry sandstone reading 0·0055 and damp sandstone from 0·0064 to 0·0085 (Everett, *loc. cit.*). The mean of these figures and a gradient of 30 m. would account for a heat flow of $2{\cdot}5 \times 10^{-6}$ cal. per sec. per sq. cm. The conductivity of the damp sandstone may be excessive, but as we have taken but little account of the rise of gradient downwards it is plain that the thermal conditions arising out of a crust uniformly radioactive and at its base maintained at a temperature in or about 1150°, is not discordant with the indications of surface gradients.

We shall now turn to the question of the thermal stability of the greater compensations. I shall take as an extreme case that of the Tibetan Plateau. Its average height over sea-level is stated to be 15,000 feet or 4575 m. Its height reckoned from the average continental surface level of 700 m.

* See Daly, Ann. J. Sc. May 1923.

is, therefore, 3875 m. Using the factor arising out of the density ratio as given above, *i. e.* 6·5, a compensation depth of 25,187 m. is obtained. Adding the continental depth and the height of the plateau, a total depth of 61 k. is arrived at.

What will be the distribution of temperature in so great a vertical depth of continental materials? The base is, say, at 1200°. The surface is at 0°. If the rise in temperature downwards exceeds at any level 1200° then there must be downward flow of heat. It is easy to see that this condition must come about. Hence there must be some level at which the direction of heat-flow changes, and this level must be that of maximum temperature. The equation connecting basal temperature with depth is $\theta = \frac{Q}{2k} D^2$, where Q is the heat generated in unit volume in unit time, k is the conductivity and D is depth. Accordingly the intersection of two parabolic curves, one displaced 1200° in the scale of temperature, may be used to determine the solution of problems of the present sort. We find that the level of maximum temperature is closely 42 k. from the surface and 19 k. from the base; the maximum temperature being 1500°.

From this we may infer that fusion must be nearly if not actually reached in this compensation. We might, of course, have chosen data which, while defensible, would give a lower internal temperature. But even as it stands it is doubtful if under the conditions of pressure actual fluidity would exist and not rather conditions of viscosity. Nor would a certain amount of fusion in such a case necessarily confer instability. For the fused interior would be on all sides shut in by a thick wall of stable material which, at least in periods succeeding revolution,—such as we now live in—must be but little hotter than the melting-point of the basalt and which being highly siliceous is, probably, mainly composed of quartz; the initial melting-point of which is 1600°.

However, even if in this particular case vertical forces could be transmitted through such a mass, it is certain that much deeper compensations would, on our data, be unstable. It is not improbable that herein we find a limit to isostatic compensation. It is stated that the Himalayas are only 80 per cent. compensated. Mountain elevation on the surface of the Globe would find limits from this cause. For observations appear to show that the rigidity of the crust is not able to carry the greater ranges without the support of compensations.

The whole matter is instructive. For the genesis of batholithic invasion from beneath, attending the rising mountain ranges,—which is one of the most striking and eloquent features of mountain structure—seems to find explanation in these simple calculations.

We may go even further. We saw above that the continental crust, taken at 35 k. thick, attains a radioactive basal temperature nearly the same as that of the substratum. Certain facts of petrological science possibly indicate that such a basal temperature would be greater than that of the magma. I refer to the evidence we possess that juvenile gases (notably water) contained in the abyssal magma might confer upon it a melting temperature somewhat lower than that which we observe in extravasated materials. If this be the case heat would flow downwards during the long era of general thermal accumulation throughout the substratum.

Now there must be a limit to such heat supplies to the magma; for under the continental base there is no escape for the heat till general fluidity is reached and the inevitable tidal movements distribute the heat into suboceanic regions. The superheated materials gravitating upwards must therefore accumulate beneath the continents, and if the temperature rises sufficiently, the base of the continental layer must melt; and when circulation and surface drift of the upper crust commence these melted materials must be carried from beneath the continents and float upwards around their margin.

These inferences seem to show that in the thermal conditions arising out of radioactive heating there arises a limit to continental thickness. Now this also controls the horizontal area of the continents, and, dependently, the area and depth of the ocean. For we see that if, originally, these lighter materials, rising like a scum to the surface of the magma, had been piled up deeper than they now are, either fortuitously or from forces originating in the rotational motion of the earth, they must inevitably have melted away beneath, until they attained their present thickness and surface extension.

June 1st, 1923.

XVIII. *On the Quantum Theory of the Complex Zeeman Effect.*

By A. M. MOSHARRAFA, *Ph.D.**

[Part of Thesis approved for the Degree of Doctor of Philosophy in the University of London.]

§ 1. *Introduction and Aim.*

IN a previous paper †, a theory of the Zeeman Effect, for hydrogen, was developed from the standpoint of the Quantum Theory. In order to be able to extend the theory to other elements, a knowledge of the atomic fields for such elements would appear to be essential. Now, an exact theory of the atomic and molecular structures of the various elements may justly be described as the goal of Spectroscopic research. And, so far, it has only been possible to present such an exact theory in the case of the simple hydrogen atom. Nevertheless, the general theory developed by Sommerfeld ‡ for what he describes as the "non-hydrogen-like" elements, although it has no claim to absolute quantitative accuracy, yet leaves little room for doubt as to the relevance of its essential qualitative features. The aim of this paper is to show that a theory of the complex Zeeman Effect is possible which assumes the essential features of Sommerfeld's theory of the atomic structure of the elements.

§ 2. *The Atomic Field.*

We assume with Sommerfeld § that the potential energy of an electron situated in the atomic field can be represented during the stationary states of the atom by

$$E_{\text{pot.}} = -\frac{eE}{r} + \frac{c_1}{r^3} + \frac{c_2}{r^5}, \quad . \quad . \quad . \quad . \quad (1)$$

where $c_1, c_2, \ldots$ are to be considered as small quantities of the 1st, 2nd, ... orders respectively, r being the distance of the electron [charge $= (-e)$] from the nucleus of the atom which is taken as origin. There is no doubt that the origin of the terms in $c_1, c_2, \ldots$ is to be sought for in the electric forces between the electron in question and the other electrified particles entering into the structure of the atom.

* Communicated by the Author.

† Roy. Soc. Proc., Series A, Feb. 1923.

‡ See A. Sommerfeld, 'Atombau und Spektrallinien,' II. Auf. Braunschweig, 1921, 4 Kap. § 6, p. 275.

§ *Loc. cit.*

And Sommerfeld * has actually expressed these quantities approximately in terms of the specific conditions of a hypothetical ring of electrons (the "inner ring"), in which a number of electrons are assumed to be in uniform rotation round the nucleus, the radius of the ring being much smaller than r. He also writes $E = ke =$ the total positive charge on the system (nucleus + "inner ring"), so that k is a whole number. Now, as Sommerfeld † himself points out, no account is taken in his theory of the reaction of the "outer" electron on the "inner ring." And Landé ‡ has shown, in the case of helium, that if such account be taken, then E is no longer an integral multiple of e, but is given by

$$E = ke(1+\kappa), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where ke has the same meaning as before, and κ is a factor depending on the type of motion of the atom.

The quantities κ, c_1, c_2, are accordingly here assumed in general to depend on the type of motion of the atom and to vary with it. We should, therefore, expect these quantities to involve not only the quantum numbers of the "outer" or, as we shall here call it, the Series electron, but also other "hidden" quantum numbers belonging to the rest of the atom. In discussing the steady states of the system, these quantities may, however, be legitimately regarded as mere constants.

It is important to note that we do not involve for the validity of our analysis any specific assumptions as to the origin of the correction terms in κ, c_1, c_2, ... etc. (such as Sommerfeld's "inner ring" hypothesis). For it is sufficient for our purpose:

(i.) that the radiation from the atom in passing from one steady state to another may be regarded as corresponding to the change in energy of a single constituent electron, and

(ii.) that the potential energy of this electron in the steady states may be considered to be given by equation (1) above.

These two hypotheses, which in effect form the essential features of Sommerfeld's analysis, are justified in some measure by the remarkable way in which they lead to the

* *Loc. cit.* p. 506.

† *Loc. cit.* p. 279.

‡ A. Landé, *Physik. Zeitschr.* xx. p. 228 (1919), and xxi. p. 114 (1920).

Rydberg and Ritz forms of the Series term respectively as successive approximations.

With regard to the effect of the magnetic forces, arising from the motion of the rest of the atom, on the series electron, this may at once be dismissed as negligible. Such an effect would, of course, lead to a splitting of the spectral lines in a similar manner to the effect of an external magnetic field. The atomic magnetic field is, however, for all ordinary elements of the order $h = \frac{ew}{ac} (=) 10^3$ Gauss, where w and a are comparable with the orbital velocity and atomic radius respectively. This would lead to a division of the lines of the order $\Delta\lambda = \frac{e\lambda^2 h}{4\pi m_0 c^2} (=) 10^{-3}$ Å in the visible portion of the spectrum, where m_0 is the mass of the electron and Å stands for Ångström units. The effect is thus too small to be detected, and we may safely dismiss it from our theory.

The effect of the change in the mass of the electron with velocity, on the other hand, is quite measurable. And although it does not play any special part in the Zeeman effect, we have nevertheless taken account of it in the analysis for the sake of completeness.

§ 3. *Application of the Method of Separation of the Variables.*

The analysis in this section is an extended form of our analysis in the last paper *, the two main new features being: the existence of the two terms in c_1 and c_2 in the expression for the energy, and the possible variation of all the three quantities E, c_1, c_2 with the external magnetic field H. The calculations are carried out to the second order of small quantities, c_1 and c_2 being considered of the 1st and 2nd orders respectively †. Powers of H beyond the first and products of H and second or higher order terms are neglected. The relativity correction terms are only taken account of to the first power in $\frac{1}{c^2}$ ‡, and their products with H or with second order terms are neglected.

Spherical polar coordinates (r, θ, ψ) are used with the

* *Loc. cit.*

† *Cf.* the beginning of § 2.

‡ c is used throughout this paper to denote the velocity of light *in vacuo*; it is to be distinguished from the coefficients c_1 and c_2.

origin in the nucleus and Oz in the direction of H. The notation is, as far as possible, kept the same as Sommerfeld's for facility of comparison.

Let W be the total energy of the series electron in one of the steady states of the atom, we have

$$W = E_{kin.} + E_{pot.} = m_0c^2\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) - \frac{eE}{r} + \frac{c_1}{r^3} + \frac{c_2}{r^5},$$

where $c\beta$ is the velocity of the electron. This may be written

$$\frac{1}{\sqrt{(1-\beta^2)}} = 1 + \frac{W + \frac{eE}{r} - \frac{c_1}{r^3} - \frac{c_2}{r^5}}{m_0c^2}. \quad . \quad . \quad (3)$$

We also have

$$\beta^2 = \frac{1}{c^2}(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\psi}^2)$$

$$= \frac{1}{c^2m^2}\left(p_r^{\,2} + \frac{1}{r^2}p_\theta^{\,2} + \frac{1}{r^2\sin^2\theta}p_\psi^{\,2}\right),$$

where

$$\left.\begin{array}{l} p_r = m\dot{r}, \quad p_\theta = mr^2\dot{\theta}, \quad p_\psi = mr^2\sin^2\theta\dot{\psi}, \\ \qquad m = m_0/\sqrt{(1-\beta^2)}, \end{array}\right\} \quad . \quad (4)$$

so that

$$\frac{\beta^2}{1-\beta^2} = \frac{1}{c^2m_0^{\,2}}\left(p_r^{\,2} + \frac{1}{r^2}p_\theta^{\,2} + \frac{1}{r^2\sin^2\theta}p_\psi^{\,2}\right),$$

or

$$\frac{1}{1-\beta^2} = 1 + \frac{1}{c^2m_0^{\,2}}\left(p_r^{\,2} + \frac{1}{r^2}p_\theta^{\,2} + \frac{1}{r^2\sin^2\theta}p_\psi^{\,2}\right). \quad (5)$$

From (3) and (5), we have

$$p_r^{\,2} + \frac{1}{r^2}p_\theta^{\,2} + \frac{1}{r^2\sin^2\theta}p_\psi^{\,2} = 2m_0W + \frac{2m_0eE}{r} - \frac{2m_0c_1}{r^3}$$

$$- \frac{2m_0c_2}{r^5} + \frac{1}{c^2}\left(W + \frac{eE}{r} - \frac{c_1}{r^3} - \frac{c_2}{r^5}\right)^2, \quad (6)$$

which defines the total energy of the motion in terms of the Hamiltonian coordinates. Now

$$\frac{dp_i}{dt} + \frac{\partial W}{\partial q_i} = -e\sum_s\left(\frac{\partial \mathbf{A}_s}{\partial q_i} - \frac{\partial \mathbf{A}_i}{\partial q_s}\right)\frac{dq_s}{dt},$$

or

$$\frac{d(p_i - e\mathbf{A}_i)}{dt} + \frac{\partial W}{\partial q_i} = -e\sum_s\frac{\partial \mathbf{A}_i}{\partial q_i}\frac{dq_s}{dt}, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where **A** is the generalized vector potential given by

$$\left(\frac{\partial \mathbf{A}_3}{\partial q_2} - \frac{\partial \mathbf{A}_2}{\partial q_3}\right) = g_2 g_3 \frac{\mathrm{H}_1}{c} \text{ and two similar equations,} \quad (8)$$

subject to the condition

$$\sum_i \frac{1}{g_i} \frac{d}{\partial q}\left(\frac{\mathbf{A}_i}{g_i}\right) + \frac{1}{c^2}\dot{\Phi} = 0, \quad . \quad . \quad . \quad (8\,a)$$

where Φ is the scalar potential and $g\,dq$ is lineal.

On applying (7) to the coordinate ψ, we get

$$\frac{d(p_\psi - e\mathbf{A}_\psi)}{dt} + 0 = 0,$$

so that

$$p_\psi - e\mathbf{A}_\psi = \text{constant} = \mathrm{F} \text{ (say)}. \quad . \quad . \quad . \quad (9)$$

Or, since

$$e\mathbf{A}_\psi = \frac{e\mathrm{H}r^2 \sin^2\theta}{2c} = m_0 w r^2 \sin^2\theta, \quad . \quad . \quad (10)$$

where

$$w = \frac{1}{2}\frac{e\mathrm{H}}{m_0 c}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

we have from (9) and (10)

$$p_\psi = \mathrm{F} + m_0 w r^2 \sin^2\theta. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Substituting from (12) in (6), we have, on multiplying all the terms by r^2 and rearranging them,

$$r^2\Bigg[-p_r^2 + 2m_0\mathrm{W} + \frac{2m_0 e\mathrm{E}}{r} - \frac{2m_0 c_1}{r^3} - \frac{2m_0 c_2}{r^5} + \frac{1}{c^2}\left(\mathrm{W} + \frac{e\mathrm{E}}{r} - \frac{c_1}{r^3} - \frac{c_2}{r^5}\right)^2 - 2m_0\mathrm{F}w\Bigg]$$
$$= p_\theta^2 + \frac{\mathrm{F}^2}{\sin^2\theta} + m_0^2 w^2 r^4 \sin^2\theta. \quad (13)$$

And in this last equation the variables are separable if we neglect the term in w^2, giving

$$r^2\Bigg[-p_r^2 + 2m_0\mathrm{W} + \frac{2m_0 e\mathrm{E}}{r} - \frac{2m_0 c_1}{r^3} - \frac{2m_0 c_2}{r^5} + \frac{1}{c^2}\left(\mathrm{W} + \frac{e\mathrm{E}}{r} - \frac{c_1}{r^3} - \frac{c_2}{r^5}\right)^2 - 2m_0\mathrm{F}w\Bigg]$$
$$= p_\theta^2 + \frac{\mathrm{F}^2}{\sin^2\theta} = p^2 \text{ (say)}, \quad (14)$$

giving

$$p_r = \sqrt{\left[A + \frac{2B}{r} + \frac{C}{r^2} + \frac{D_1}{r^3} + \frac{D}{r^4} + \frac{D_2}{r^5}\right]}, \quad . \quad (15)$$

$$p_\theta = \sqrt{\left(p^2 - \frac{F^2}{\sin^2\theta}\right)},$$

where

$$\left.\begin{aligned} A &= 2m_0 W\left(1 + \frac{W}{2m_0c^2} - \frac{Fw}{W}\right), \\ B &= eEm_0\left(1 + \frac{W}{m_0c^2}\right), \\ C &= -\left(p^2 - \frac{e^2E^2}{c^2}\right), \\ D_1 &= -2m_0c\left(1 + \frac{W}{m_0c^2}\right), \\ D &= -\frac{2eEc_1}{c^2}, \\ D_2 &= -2c_2m_0, \end{aligned}\right\} \quad . \quad . \quad (17)$$

omitting terms of higher orders.

We now apply Wilson's extended restrictions referred to in our last paper *, viz.

$$\int_0^\bullet (p_i - e\mathbf{A}_i)dq_i = n_ih, \qquad i = 1, 2, 3, \ldots \quad . \quad (18)$$

Thus, since

$$\mathbf{A}_r = \mathbf{A}_\theta = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

we have

$$\left.\begin{aligned} \int_0^\bullet p_r dr &= n_1h \quad (\alpha), \\ \int_0^\bullet p_\theta d\theta &= n_2h \quad (\beta), \\ \int_0 (p_\psi - e\mathbf{A}_\psi) &= n_3h \quad (\gamma). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (20)$$

And on substituting from (9), (15), and (16), we get

$$\int_0 \sqrt{\left[A + \frac{2B}{r} + \frac{C}{r^2} + \frac{D_1}{r^3} + \frac{D}{r^4} + \frac{D_2}{r^5}\right]} dr = n_1h, \quad (21)$$

$$\int_0 \sqrt{\left[p^2 - \frac{F^2}{\sin^2\theta}\right]}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

$$\int_0 F d\psi = n_3h. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

* *Loc. cit.* § 4, equation (17 *a*).

The first integral is evaluated in Appendix I. at the end of this paper. The second integral has already been evaluated in our last paper *. The third integral extends over the period of ψ, namely from 0 to 2π, and is thus at once seen to be equal to $2\pi F$. We have, using these results,

$$-2\pi i\left[\sqrt{C}-\frac{B}{\sqrt{A}}-\frac{1}{2}\frac{B}{C\sqrt{C}}\left(D_1+\frac{5}{2}D_2\frac{B^2}{C^2}+\frac{15}{8}D_1{}^2\frac{B}{C^2}\right.\right.$$
$$\left.\left.-\frac{3}{2}\frac{BD}{C}\right)+\frac{3}{4}\frac{A}{C^2\sqrt{C}}\left(D_2B+\frac{1}{4}D_1{}^2-\frac{1}{3}CD\right)\right]=n_1h, \quad (24)$$

$$p-F=\frac{n_2h}{2\pi}, \quad \ldots \ldots \quad (25)$$

$$F=\frac{n_3h}{2\pi}. \quad \ldots \ldots \quad (26)$$

And from the last two equations we have

$$p=\frac{nh}{2\pi}, \quad \ldots \ldots \quad (27)$$

where

$$n=n_2+n_3. \quad \ldots \ldots \quad (28)$$

In order to calculate the energy W of the static paths, let

$$W=W_0+\delta_w W+\delta_c W, \quad \ldots \quad (29)$$

where W is the value obtained for $w=\frac{1}{c}=0$ (*i. e.*, the value calculated in the absence of the field and without taking account of the relativity correction), $\delta_w W$ and $\delta_c W$ being first-order terms in w and $\frac{1}{c^2}$ respectively; and adopt a similar notation for the coefficients A, B, C, etc. We have from (17)

$$\left.\begin{array}{lll} A_0=2m_0W_0, & B_0=eE_0m_0, & C_0=-p^2, \\ D_{1,0}=-2m_0c_{1,0}, & D_0=0, & D_{2,0}=-2c_{2,0}m_0, \end{array}\right\} \quad (30)$$

* *Loc. cit.* § 2, Part II., equations (46) to (50).

so that we have from (24) and (27)

$$-nh+\frac{2\pi i m_0 e\mathrm{E}_0}{\sqrt{\mathrm{A}_0}}-a_0 h+\frac{\alpha_0\mathrm{A}_0}{2m_0}=n_1 h, \quad . \quad . \quad (31)$$

with the following contractions :

$$a_0=\frac{(2\pi)^4 m_0^2 e\mathrm{E}_0}{n^3h^4}\left(c_{1,0}-\frac{5}{2}c_{2,0}\frac{(2\pi)^4 m_0^2e^2\mathrm{E}_0^2}{n^4h^4}\right.$$

$$\left.+\frac{15}{4}c_{1,0}\frac{(2\pi)^4 m_0^2 e^2\mathrm{E}^2}{n^4h^4}\right) \quad . \quad (32)$$

$$=-3\frac{(2\pi)^6 m_0^3}{n^5h^5}\left(c_{2,0}e\mathrm{E}_0-\frac{1}{2}c_{1,0}^2\right), \quad . \quad . \quad . \quad . \quad . \quad (33)$$

which leads to the Ritz form for W_0 already obtained by Sommerfeld *, viz.,

$$\mathrm{W}_0=-\frac{\mathrm{N}h(\mathrm{E}_0/e)^2}{(n+n_1+a_0-\alpha_0\mathrm{W}_0/h)^2}, \quad . \quad . \quad . \quad (34)$$

where N is the Rydberg constant

$$\mathrm{N}=\frac{2\pi^2 m_0 e^4}{h^3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

With regard to the quantities E, c_1, c_2, ... we assume † a change in their values due to the change in the type of motion of the atom produced by the magnetic field. We accordingly write

$$\left.\begin{aligned}\mathrm{E}&=\mathrm{E}_0+\underset{w}{\delta}\mathrm{E},\\ c_1&=c_{1,0}+\underset{w}{\delta}c_1,\\ c_2&=c_{2,0}+\underset{w}{\delta}c_2,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (36)$$

where $\underset{w}{\delta}\mathrm{E}/\mathrm{E}_0$, $\underset{w}{\delta}c_1/c_{1,0}$, $\underset{w}{\delta}c_2/c_{2,0}$ are considered as first-order terms in w. The last of these quantities, however, will always occur in our analysis in terms already containing c_2, and may therefore be dismissed as beyond our degree of approximation. Thus, for our purpose,

$$c_2=c_{2,0}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

We can now write for the coefficients involved in (24) to

* *Loc. cit.* p. 506.
† See § 2.

our degree of approximation

$$\left.\begin{aligned}
A &= A_0\left(1+\frac{\underset{w}{\delta W}}{W_0}+\frac{W_0}{2m_0c^2}+\frac{\underset{c}{\delta W}}{W_0}-\frac{Fw}{W_0}\right),\\
B &= B_0\left(1+\frac{W_0}{m_0c^2}+\frac{\underset{w}{\delta E}}{E_0}\right),\\
C &= C_0\left(1+\frac{e^2E_0^2}{p^2c^2}\right)=C_0\left(1+\frac{e^2E_0^2}{C_0c^2}\right),\\
D_1 &= D_{1,0}\left(1+\frac{W_0}{m_0c^2}+\frac{\underset{w}{\delta c_1}}{c_{1,0}}\right),\\
D &= D_0=\frac{eED_{1,0}}{m_0c^2},\\
D_2 &= D_{2,0},
\end{aligned}\right\} \quad . \quad (38)$$

so that the values of the respective terms in (24) may now be written down

$$\left.\begin{aligned}
-2\pi i\sqrt{C} &= 2\pi i\sqrt{C_0}\left(1+\frac{e^2E_0^2}{2C_0c^2}\right),\\
\frac{2\pi iB}{\sqrt{A}} &= \frac{2\pi iB_0}{\sqrt{A_0}}\left(1+\frac{3W}{4m_0c^2}-\frac{\underset{c}{\delta W}}{2W_0}\right.\\
&\qquad\qquad\left.+\frac{\underset{w}{\delta E}}{E_0}-\frac{\underset{w}{\delta W}}{2W_0}+\frac{Fw}{2W_0}\right),\\
\frac{\pi iBD_1}{C\sqrt{C}} &= \frac{\pi iB_0D_{1,0}}{C_0\sqrt{C_0}}\left(1+\frac{2W_0}{m_0c^2}+\frac{\underset{w}{\delta E}}{E_0}+\frac{\underset{w}{\delta c_1}}{c_{1,0}}-\frac{3e^2E_0^2}{2C_0c^2}\right),\\
-\frac{3\pi iB^2D}{2C^2\sqrt{C}} &= -\frac{3\pi iB_0^2D}{2C_0^2\sqrt{C_0}}=-\frac{3\pi ieE_0B_0^2D_{1,0}}{2C_0^2\sqrt{C_0}m_0c^2},\\
\frac{\pi iAD}{2C\sqrt{C}} &= \frac{\pi iA_0D}{2C_0\sqrt{C_0}}=\frac{\pi iA_0eE_0D_{1,0}}{2m_0c^2C_0\sqrt{C_0}},
\end{aligned}\right\} (39)$$

and the rest of the terms involved in (24) are not affected to our degree of approximation, either by the magnetic field or by the relativity correction. $\underset{w}{\delta W}$ and $\underset{c}{\delta W}$ are now obtained by equating the sum of the terms in w and in $\frac{1}{c^2}$ on the

right-hand side of (39) respectively to zero. We accordingly have

$$\delta_{w} W = \delta_{w, c_1} W + \delta_{w, E} W + \delta_{w, F} W, \quad . \quad . \quad . \quad (40)$$

where

$$\left.\begin{aligned} \delta_{w, E} W &= 2W_0\left(1+\frac{D_{1,0}}{2C_0\sqrt{C_0}}\right)\frac{\delta_w E}{E_0}, \\ \delta_{w, c_1} W &= \frac{W_0 D_{1,0}\sqrt{A_0}}{C_0\sqrt{C_0}}\frac{\delta_w c_1}{c_{1,0}}, \\ \delta_{w,F} W &= Fw, \end{aligned}\right\} \quad . \quad . \quad (41)$$

which yield on substituting from (30) and remembering that $\sqrt{C_0}$ must be taken as the negative (imaginary) root of C_0:

$$\left.\begin{aligned} \delta_{w, E} W &= 2W_0\left(1+\frac{im_0\sqrt{2m_0W_0}}{p}c_{1,0}\right)\frac{\delta_w E}{E_0}, \\ \delta_{w, c_1} W &= \frac{2im_0W_0\sqrt{2m_0W_0}}{p^3}c_{1,0}\frac{\delta_w c_1}{c_{1,0}}, \\ \delta_{w, F} W &= Fw. \end{aligned}\right\} \quad . \quad (42)$$

And similarly we have for $\delta_c W$ from (39) and (30) after some reduction

$$\delta_c W = \frac{2W_0\sqrt{2m_0W_0}}{m_0c^2}\left[\frac{3W_0}{4\sqrt{2m_0W_0}} - \frac{ieE_0}{2p} + \frac{3im_0W_0}{p^3}c_{1,0} + \frac{3ie^2E_0{}^2m_0c_{1,0}}{p^5}\right]. \quad (43)$$

Or, if we substitute the values of W_0 and p from (34) and (27) respectively and introduce a_0 as defined by (32), we have from (42) and (43) to our degree of approximation,

$$\left.\begin{aligned} \delta_{w, E} W &= -\frac{2N_0'h}{(n+n_1)^2}\left(1-\frac{3a_0}{n+n_1}\right)\frac{\delta_w E}{E_0}, \\ \delta_{w, c_1} W &= \frac{2N_0'ha_0}{(n_1+n)^3}\frac{\delta_w c_1}{c_{1,0}}, \\ \delta_{w, F} W &= n_3\frac{wh}{2\pi}, \end{aligned}\right\} \quad . \quad . \quad (44)$$

$$\underset{e}{\delta} W = -\frac{\gamma N_0' h}{(n+n_1)^4}\left[\frac{1}{4}(1-a_0 Z)+\frac{n_1}{n}\left(1-\frac{3a_0}{n+n_1}\right)\right], \quad (45)$$

where

$$N_0' = N \times \left(\frac{E_0}{e}\right)^2, \quad (46)$$

$$\gamma = \frac{(2\pi)^2 e^2 E_0^2}{c^2 h^2}, \quad (47)$$

$$Z = 12(n_1+n)\left(\frac{2}{n^2}-\frac{1}{(n_1+n)^2}\right). \quad (48)$$

§ 4. *Significance of the Results of the Previous Section.*

We see from (44) that the exact value of W, and therefore the corresponding change in frequency $\underset{w}{\delta}\nu$, depends on the two variational terms $\underset{w}{\delta} E/E_0$ and $\underset{w}{\delta} c_1/c_{1,0}$ as well as on a_0. The last quantity as given by equation (32) is identified in Sommerfeld's theory * with the corresponding experimental quantity involved in the Ritz form of the Series term [see equation (34)], where $n=1, 2, 3, 4$ ... are assumed to correspond to S, P, D, B ... terms respectively, *i. e.* to the variable terms in the II. Subordinate, Principal, I. Subordinate, Bergman, ... series. When, however, the values of the universal constants are substituted in (32), no satisfactory numerical agreement with experimental data is obtained, if we consider E_0, $c_{1,0}$, ... as absolute constants. It is not our object in this paper to attempt an elucidation of this point. But we should like to point out that if E_0, $c_{1,0}$, ... are regarded not as absolute constants but as depending on the type of motion of the atom (see § 2) and therefore involving unknown quantum numbers, and possibly n_1, n_2, n_3 as well, then it would seem possible to attempt an explanation, not only of the numerical discrepancies in the values of a_0, but also perhaps of the doublet and triplet structures of the series †. This would, of course, involve a more or less exact

* *Loc. cit.*

† No satisfactory explanation has so far been put forward of this fine structure. An additional atomic field arising from an electric doublet of charge e and moment M along Oz would lead to a fine structure expression,

$$\delta W = \frac{24\pi^4 m_0^2 e^2 M N_0'}{h^3 n^3 (n_1+n)^3}\left(1-\frac{n_3^2}{n^2}\right)^2,$$

which, however, does not give satisfactory agreement with observation (see a paper by F. Tank in the *Ann. d. Phys.* lix. p. 293, 1919). An explanation based on the relativity correction is equally unsuccessful

determination of E_0, $c_{1,0}$, ... for the various types of atomic motions, a task for which our present theoretical equipments may not prove adequate. For our purpose here, however, it is not necessary to identify a_0 with its corresponding experimental values.

With regard to the two variational quantities $\frac{\delta_w E}{E_0}$ and $\frac{\delta_w c_1}{c_{1,0}}$, we are again led for their exact evaluation to a consideration of the motion of the whole atom and the variation of that motion with an external magnetic field. We may thus expect them to involve unknown quantum numbers as well as n_1, n_2, n_3. Now the order of magnitude, at least of those two variational quantities, is determinable from our analysis; for it must be the same as that of the other variational quantities involved (*e. g.* $\delta_{w,F} W/W_0$). And it is the characteristic feature of our analysis that by assuming small variations of the energy coefficients E_0, $c_{1,0}$, ... etc. proportional to the magnetic fields and of the same order of relative magnitude as the variations in the other constants of the original motion, we find it possible to arrive at a satisfactory explanation of the complex Zeeman Effect. We accordingly write:

$$\left.\begin{aligned} \frac{\delta_w E}{E_0} &= \frac{w(n+n_1)^2}{4\pi N_0'}\phi_1, \\ \frac{\delta_w c_1}{c_{1,0}} &= \frac{w(n_1+n)^2}{4\pi N_0'}\phi_2, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (49)$$

where ϕ_1 and ϕ_2 are functions of whole numbers.

From (40), (44), and (49) we have

$$\delta_w W = \frac{wh}{2\pi}\left((n_3-\phi_1)+a_0\times\frac{(3\phi_1+\phi_2)}{n+n_1}\right). \quad . \quad . \quad (50)$$

Now, Sommerfeld * has shown that the splitting of the lines in the most general type of the Zeeman effect could be

(see end of our §4). The view I have here ventured to put forward that the structure of the series should depend on the type of motion of the atom would appear to receive support from the fact that the same element emits lines of different structural characters (doublets, triplets, etc.) under different experimental conditions, *e. g.* for its arc and spark spectra respectively.

* *Loc. cit.* p. 537. See also a paper by the same author in the *Ann. d. Phys.* 1920.

exactly represented for any spectral line by the formula

$$\Delta\nu = \frac{w}{2\pi}\left(\frac{q_1}{r_1} - \frac{q_2}{r_2}\right),$$

where r_1 and r_2 (the "Runge denominators") depend only on the nature of the term to which they correspond (*i. e.*, whether S−, P−, D−, or B− term, etc.) and the structural character of the spectral line (*i. e.*, whether a simple line, a member of a triplet or doublet respectively), and q_1, q_2 (the "Runge numerators") assume different values for the different Zeeman components out of the series $0, \pm 1, \pm 2, \ldots, \pm r_1, \pm(r_1+1), \ldots$ and $0, \pm 1, \pm 2, \ldots, \pm r_2, \pm(r_2+1), \ldots$ respectively. Also $r_1 r_2 = r$, the Runge number, which is fixed for each type of decomposition. The values of the Runge denominators for the respective terms are given by the following table, which we copy from Sommerfeld :—

Table A.

	S.	P.	D.	B.
Simple Lines	1	1	1	1
Triplet Lines	1	2	3	(4)
Doublet Lines	1	3	5	(7)

The table is based on the experimental results of Paschen and Back except for the two bracketed numbers, for which no corresponding experimental data are so far available and which have been written down by analogy. If we remember that the S, P, D, B terms etc. are characterized by the azimuthal quantum number $n=1, 2, 3, 4$, etc., we at once see from the table that the Runge denominator is numerically equal to 1 for simple lines, n for triplet lines, and $2n-1$ for doublet lines.

We may therefore identify it with these numbers * for the respective cases. And in order to arrive at an exact and full representation of the frequencies in the most general type of Zeeman decomposition, it is only necessary now to write from (50)

$$\left.\begin{aligned} n_3 - \phi_1 + \frac{a_0}{n+n_1}(3\phi_1+\phi_2) &= \frac{n_3}{1} && \text{for simple lines,} \\ \text{or} \quad &= \frac{\phi}{n} && \text{,, triplets,} \\ \text{or} \quad &= \frac{\phi'}{2n-1} && \text{,, doublets,} \end{aligned}\right\} \quad (51)$$

* The Runge denominators may not be identical with these numbers—*e. g.*, they may be given by functions of unknown quantum numbers, which functions must, however, be numerically equal to 1, n, and $2n-1$ for the respective cases.

where ϕ and ϕ' assume values out of the series:

$$0,\ \pm 1,\ \pm 2,\ \ldots,\ \pm n,\ \pm(n+1),\ \ldots \text{ etc.} \quad \text{and}$$
$$0,\ \pm 1,\ \pm 2,\ \ldots,\ \pm(2n-1),\ \pm 2n,\ \ldots \text{ etc.} \quad \text{respectively.}$$

From (51) we have as a first approximation for ϕ_1, neglecting terms in a_0:

$$\left.\begin{aligned} [\phi_1]_0 &= 0 && \text{for simple lines,} \\ \text{or } &= n_3 - \frac{\phi}{n} && \text{,, triplets,} \\ \text{or } &= n_3 - \frac{\phi'}{2n-1} && \text{,, doublets.} \end{aligned}\right\} \quad . \quad (52)$$

Or, if we take account of terms in a_0 to the first order only in (51), we may write

$$\left.\begin{aligned} \phi_1 &= [\phi_1]_0 + \frac{a_0\sigma_i}{n+n_1}, \\ [\phi_2]_0 &= \sigma_i - 3[\phi_1]_0, \end{aligned}\right\} \ i = 1, 2, 3, \quad . \quad . \quad (53)$$

where the coefficients σ_1, σ_2, σ_3 correspond to simple lines, triplets, or doublets respectively, but cannot be determined by our analysis, and the notation $[\ ]_0$ denotes values of the quantities involved calculated for $c_{1,0} = c_{2,0} = \ldots = 0$.

The two functions, ϕ and ϕ', introduced in (51) must be understood to involve "hidden" quantum numbers arising from a quantization of the general motion of the whole atom. The intensities and types of polarization of the various Zeeman components will depend on the nature of these two functions, and may be determinable by the application of Bohr's Principle of Correlation to the internal mechanism of the atom. Such an application would be equivalent to a form of Choice Principle relating to the "hidden" quantum numbers, and should lead to an explanation not only of the questions of intensity and polarization, but also of the way in which ϕ and ϕ' assume different groups of values for the different components of a doublet or a triplet.

With regard to the effect of the relativity correction on the appearance of the lines, we see from (45), (46), (47), and (48) that if we assume n to be fixed for each series term, then, since $(n+n_1)$ is also fixed, it follows that n_1 is also fixed, and hence δW is a one-valued quantity. Thus the effect of the

c

relativity correction on any spectral line would on Sommerfeld's theory * consist in a slight modification in the position of the line given by

$$h\delta\nu = \underset{c}{\delta} W(m) - \underset{c}{\delta} W(n), \quad . \quad . \quad . \quad . \quad (54)$$

where $\underset{c}{\delta} W(m)$ and $\underset{c}{\delta} W(n)$ are given by (45) for the quantum numbers (m, m_1) and (n, n_1) respectively. This leads to no fine structure whatever, and an explanation of the doublet and triplet structure of the lines must be sought for elsewhere.

§ 5. *Conclusion.*

In his remarks on the array of figures given in Table A of the last article, Prof. Sommerfeld † writes : "only so much appears to be certain : that the harmony of integral numbers represented by our Runge Denominators has its ultimate foundation in the rules of hidden quantum numbers, and possesses a quantum derivation." We have attempted in this paper to inquire into the nature of this derivation, and have shown how it may be possible to ascribe the most complex and general types of Zeeman decomposition to small variations in the atomic field arising from the establishment of the external magnetic field. Our theory is, however, limited by our very deficient knowledge of the exact nature of the atomic field itself. The quantities E_0, $c_{1,0}$, $c_{2,0}$, ..., etc. and the way in which they arise for the different types of atomic motions being only partly comprehended at present, it does not seem yet promising to attempt an evaluation of their variations with a magnetic field from first principles. Thus we have to satisfy ourselves in § 4 with the assumption expressed by equations (51), which may be regarded as empirical. This assumption may, on the other hand, be looked upon as a guide to the nature of the terms E_0, $c_{1,0}$, ..., etc. themselves, and may eventually serve as a fresh test for possible theories which claim to assign exact values to these quantities from first principles. Thus the phenomenon of the Zeeman Effect may on the Quantum Theory, as it did on the classical theory, throw light on the problems of atomic structure and radiation.

* *Loc. cit.*

† *Loc. cit.* p. 542 ; this is a free translation from the original, which is in German.

APPENDIX I. (TO § 3).

To evaluate the integral

$$J_1=\int_0^{\infty}\sqrt{\left[A_0+\frac{2B}{r}+\frac{C}{r^2}+\frac{D_1}{r^3}+\frac{D}{r^4}+\frac{D_2}{r^5}\right]}\,dr. \qquad \text{(i.)}$$

This differs from the similar integral evaluated by Sommerfeld (‘Atombau, u.s.w.’ p. 476, under (e)) by the appearance of the term in D. We have to our degree of approximation

$$J_1=J_2+\tfrac{1}{2}D_1J_3+\tfrac{1}{2}DJ_4+\tfrac{1}{2}D_2J_5-\tfrac{1}{8}D_1^2J_6, \quad . \quad . \qquad \text{(ii.)}$$

where

$$J_2=\int_0^{\infty}\sqrt{\left[A+\frac{2B}{r}+\frac{C}{r^2}\right]}\,dr=-2\pi i\left(\sqrt{C}-\frac{B}{\sqrt{A}}\right),$$

$$J_3=\int_0^{\infty}\left(A+\frac{2B}{r}+\frac{C}{r^2}\right)^{-1/2}\frac{dr}{r^3}=+2\pi i\,\frac{B}{C\sqrt{C}},$$

$$J_4=\int_0^{\infty}\left(A+\frac{2B}{r}+\frac{C}{r^2}\right)^{-1/2}\frac{dr}{r^4},$$

$$J_5=\int_0^{\infty}\left(A+\frac{2B}{r}+\frac{C}{r^2}\right)^{-1/2}\frac{dr}{r^5}=-\frac{2\pi i}{2\sqrt{C}}\left(\frac{3AB}{C^2}-5\frac{B^3}{C^3}\right),$$

$$J_6=\int_0^{\infty}\left(A+\frac{2B}{r}+\frac{C}{r^2}\right)^{-3/2}\frac{dr}{r^6}=+\frac{3\pi i}{C\sqrt{C}}\left(\frac{A}{C}-5\frac{B^2}{C^2}\right),$$

the values of all integrals except J_4 being obtained from Sommerfeld's work, already referred to. To evaluate J_4 we note that it is regular at infinity, so that

$$\begin{aligned}J_4&=\int_0^{\infty}\frac{dr}{r^3}(C+2Br+Ar^2)^{-1/2}\\&=\frac{1}{\sqrt{C}}\int_0^{\infty}\frac{dr}{r^3}\left[1-\left(\frac{B}{C}r+\frac{A}{2C}r^2\right)+\frac{3}{2}\left(\frac{B}{C}r+\frac{A}{2C}r^2\right)^2+\ldots\right]\\&=-\frac{2\pi i}{\sqrt{C}}\left[\frac{3B^2}{2C^2}-\frac{A}{2C}\right];\end{aligned}$$

and we finally have from (ii.)

$$\begin{aligned}J_1=-2\pi i\Bigg\{&\sqrt{C}-\frac{B}{\sqrt{A}}\\&-\frac{1}{2}\frac{B}{C\sqrt{C}}\left(D_1+\frac{5}{2}D_2\frac{B^2}{C^2}+\frac{15}{8}D_1^2\frac{B}{C_2}-\frac{3}{2}\frac{BD}{C}\right)\\&+\frac{3}{4}\frac{A}{C^2\sqrt{C}}\left(D_2B+\frac{1}{4}D_1^2-\frac{1}{3}CD\right)\Bigg\}, \quad . \qquad \text{(iii.)}\end{aligned}$$

$\sqrt{C}$ being (as pointed out by Sommerfeld) taken as the negative (imaginary) root of C. This is the result quoted in the text.

King's College, London,
April 1922.

XIX. *The Auroral Spectrum and the Upper Strata of the Atmosphere. Preliminary Communication.* *By* L. VEGARD, *Doctor of Science, Professor of Physics at the University of Christiania* *.

DURING recent years I have been investigating the auroral spectrum. In the winter 1912–13, I undertook an expedition to Bossekop in Finmarken †, the main object of which was to study the auroral spectrum. With a spectrograph which combined a considerable dispersion with a great power of light, I succeeded in photographing a few of the strongest lines in the blue part of the auroral spectrum, and it was proved that these lines belong to the so-called negative band spectrum of nitrogen.

During the winter of 1921, I continued investigations in Christiania, and here I made determinations of the green auroral line ‡. As the result of a number of measurements, I found the auroral line to have the following wave-length :

$$\underline{\lambda = 5578{\cdot}0} \text{ (intern. units).}$$

Although the auroral line was determined with such an accuracy that the error is only a fraction of an Å unit, the origin of the line remained as mysterious as ever.

It was to be hoped that a more complete investigation of the whole auroral spectrum might give us some valuable information also with regard to the origin of the green line. But apart from this question, the determination of the auroral spectrum is a problem of the very greatest importance, on account of its bearing on the question regarding the constitution of the upper strata of the atmosphere and the nature of the cosmic electric rays producing the aurora borealis.

In the year 1921 the "Government Fund for Scientific Research" furnished me with the necessary means for taking up this work in a more systematic way. A more complete description of experimental arrangement will be given in a later work. Presently I shall only mention that during the last winter (1922–23) three spectrographs, which were put up last summer, have been at work at the Geophysical Institute of Tromsø, where the top roof of the building has

* Communicated by the Author.

† L. Vegard, *Phys. Z. S.* xiv. p. 677 (1913); *Ann. d. Phys.* l. p. 853 (1916); Bericht über eine Expedition nach Finmarke, *Christiania Vid. selsk. skr. Mat. kl.* 1916, No. 7.

‡ L. Vegard, *Geofysiske Publikationer*, vol. ii. No. 5.

kindly been put at my disposal by the Director, O. Krogness. For the sake of convenience we shall give the spectrographs the following notation, viz. I., II., and III. (Roman numerals), where

I. is a quartz spectrograph with a fairly large dispersion and of high light power, for studying the ultra-violet part of the spectrum ;

II. is a fairly big glass spectrograph with a considerable dispersion and a fairly high light power ;

III. is a small glass spectrograph with the largest possible light power, but with a much smaller dispersion.

The spectrographs I. and II. were specially designed for an accurate determination of the wave-length of the lines that might appear in the auroral spectrum during the time of exposure, which had to be very long.

The small glass spectrograph, III., was constructed for the study of possible variations of the auroral spectrum, and also to learn how many lines could be observed in the visible part of the spectrum.

The big spectrographs I. and II. were mounted in a wooden box where the temperature could be regulated automatically. The whole box could be turned about a horizontal and a vertical axis.

These spectrographs have been in operation during the last winter (1922–23), and in this work I have been very ably assisted by Mr. Einar Tonsberg. We have already obtained a number of plates showing a considerable number of lines. The plates have been measured at the Physical Institute of Christiania, and in this work I have been ably assisted by Mr. Jonathan Aars.

With the quartz spectrograph we have obtained three spectrograms taken on "Imperial Eclipse" plates. The time of exposure and the lines obtained are given in Table I.

It appears that with a time of exposure of 15–20 hours of northlight a considerable number of lines can be obtained on the plate by means of the quartz spectrograph. It should be noticed that in the spectra 1 and 2 the strongest lines, such as 4278 and 3914, are over-exposed.

To give an idea of the relative strength of the lines, I have given them intensities from 1–10. Later on, I intend to give more accurate intensity values based on quantitative measurements with a registering micro-photometer.

On the spectrograms from the quartz spectrograph 21 lines have been measured.

TABLE I.

Spectrograph I. (quartz spectrograph).

Photographic plates: "Imperial Eclipse."

Plate number. 1.	2.	3.	Mean.	Intensity.
Time of exposure in hours. 15·4	18·7	4		
λ.	λ.	λ.	λ.	
3134·4	3135·6	......	3135·0	3
3160·4	3159·7	......	3160·0	5
3208·3	......	......	3208·3	1
3284·9	......	......	3284·9	1
3371·5	3371·1	3371·2	3371·3	6
3432·7	(3431·2)	......	3432·7	2
3467·8	(3466·7)	......	3467·8	2
3502·9	......	......	3502·9	1
3535·7	3537·2	3535·5	3536·1	5
3575·6	3577·2	3577·8	3576·9	6
3710·7	3711·6	......	3711·1	3
3756·3	3755·1	......	3755·7	5
3805·2	3805·5	......	3805·4	5
3911·6	3915·3	3914·5	3913·8	10
3997·5	3998·5	......	3998·0	4
4054·3	4058·1	......	4056·2	4
4239·1	4236·6	......	4237·8	4
4277·3	4278·2	4277·6	4277·7	8
4421·5	......	......	4421·5	2
4653·1	4652·3	......	4652·7	3
4708·6	4706·7	......	4707·7	4

The spark spectrum of Cd used for comparison.

With the big glass spectrograph we at first used an Imperial panchromatic plate B. This plate was developed after an exposure of 20 hours of northlight, but the plate showed only the green auroral line. The He spectrum was used for comparison. The measurements gave for the auroral line the following wave-length:

$$\underline{\lambda = 5578 \cdot 2.}$$

Afterwards an Imperial Eclipse plate was put in, and developed after 18 hours of exposure. Only the six strongest lines in blue and violet were visible. The He comparison spectrum was over-exposed, and, in consequence, the accuracy of the wave-length determinination may be somewhat reduced.

The measurements gave the following results :

λ.
4708·7 Å
4651·1 „
4277·9 „
4266·8 „
4236·3 „
3913·3 „

The last line was very faint, so it could hardly be seen in the microscope. These are mainly the same lines as those which I observed in the year 1912–13. There is only the difference that now I get a faint diffuse line at about 4266·8, while in 1912–13 this line was not distinctly seen on the plate. A faint line, 4200, was just noticeable on one of my plates from the year 1912–13.

With the small glass spectrograph, III., the strongest lines (even with panchromatic plates) could be obtained with a time of exposure of 30–60 minutes.

Table II.

Spectrograph III. (small glass spectrograph).

Plate.						Mean.	Intensity.	Plate 17.
3 *a.*	3 *c.*	3 *d.*	4.	6.	7.			
Time of exposure in hours.								
0·8	0·8	0·5	1·6	1·4	0·8			6·8
...	[3745·5]	[3751·3]	...	...	...	...	1	3758·5
...	3810·6	3804·9	...	3800 8	...	3805·4	1	3807·0
3914·6	3913·3	...	[3908·3]	...	...	3913·9	10	3913·7
..	...	...	...	...	...	...	2	3941·5
...	3997·4	3999·1	...	3998·1	[4005·2]	3998·2	2	4000·4
...	4057·8	4058·4	...	4054·3	4051·9	4055·6	1	4059·7
...	...	...	...	...	...	...	1	4182·5
...	...	...	...	...	...	...	1	4200·0
4231·4	.239·1	4236·8	...	4235·8	4241·5	4236·9	5	4238·4
4278·5	4276·4	...	[4269·8]	...	...	4277·4	9	4279·0
...	...	...	...	4331·3	...	..	2	4345·8
...	...	...	...	...	...	...	1	4378·9
...	...	...	...	4424·7	...	...	2	4426·5
...	...	...	...	...	...	...	1	4478·5
...	...	...	...	...	...	...	1	4552·1
...	...	...	...	...	...	...	1	4591·9
...	4655·4	4653·7	...	4651·6	4651·4	4653·0	3	4650·8
4707·0	4713·3	...	...	...	...	4710·1	5	4708·3
...	...	...	...	...	...	...	1	4779·2
...	...	..	5582·8	...	5574·0	5578·4	1	4857·4

The spark of Cd used for comparison.

A considerable number of exposures have been made with this spectrograph, but only a few have been found sufficiently good for measurements, many of them showing only the same strongest lines that were more accurately determined with the two big spectrographs. But in order to see how many lines could be expected in the blue and violet part of the spectrum, a plate was taken (no. 17) which was exposed for 6·8 hours.

In Table II. are given the lines measured on some of the plates with a fairly small time of exposure. The last column contains the lines measured on the plate no. 17. As we can see, 20 lines have been measured, and probably some more lines can be traced by means of the micro-photometer.

An estimate of the intensities as they appeared on the plate no. 17 is also given in the table.

Identification of the Lines.

The observations carried out in the year 1912–13 gave the result that the strongest lines in the blue part of the auroral spectrum were due to nitrogen and belonged to the so-called negative band spectrum.

Comparing the new lines found with those of the N spectrum, we at once see that nearly all of them must be ascribed to nitrogen.

On the left side of Table III. have been entered the auroral lines observed, and on the right side the corresponding N lines are given. In a column headed "Type" is given the type of spectrum to which the line belongs. The numbers and the classification are taken from H. Kayser, *Handbuch der Spectroscopie*. N.B. and P.B. mean negative and positive bands ; L.S. means line spectrum.

All stronger lines in the ultra-violet found by means of the quartz spectrograph are quite accurately determined, and the errors less than 1 Å unit. The agreement between the auroral lines and the corresponding N lines is a very close one throughout the ultra-violet part—also the stronger lines of the visible part of the spectrum are quite accurately determined, and the identification quite certain.

The lines which are only determined by means of the small spectrograph may be somewhat less accurate, and there may be errors in the wave-length of 2–3 Å, but still I think the identification also of these lines is pretty certain.

TABLE III.

Auroral lines observed with spectrograph.			Lines of nitrogen spectrum.		
I.	II.	III.	λ.	Type.	Observer.
3165·0 Å	...	...	3135·9 Å	P.B.	Hermesdorf.
3160·0 „	...	...	3159·2 „	„	„
3208·3 „	...	...	?	„	„
3284·9 „	...	...	3285·3 „	„	„
3371·3 „	...	...	3371·5 „	„	„
3432·7 „	...	...	?		
3467·8 „	...	...	3468·1 „	„	Deslandres.
3502·9 „	...	...	3500·5 „	„	Hermesdorf.
3536·1 „	...	...	3536·8 „	„	„
3576·9 „	...	...	3577·0 „	„	„
3711·1 „	...	...	3710·7 „	„	„
3755·7 „	...	3758·5 Å	3755·5 „	„	„
3805·4 „	...	3807·0 „	3805·1 „	„	„
3913·8 Å	3913·3 Å	3913·7 Å	3914·4 „	N.B.	Deslandres.
...	...	3941·5 „	3943·1 „	P.B.	Hermesdorf.
3998·0 „	...	4000·4 „	3998·5 „	„	„
4056·2 „	...	4059·7 „	4058·7 „	„	Hasselberg.
...	...	4182·5 „	?		
...	...	4200·0 „	4201·0 „	„	Hasselberg. Or N.B. 4198·8.
4237·8 „	4236·3 „	4238·4 „	4236·3 „	N.B.	Hasselberg.
...	4266·8 „	...	4269·4 „	P.B.	Hasselberg. Diffuse line, perhaps shade of a band.
4277·7 „	4277·9 „	4279·0 „	4278·0 „	N.B.	Hasselberg.
...	...	4345·8 „	4343·8 „	P.B.	„
...	...	4378·9 „	4379·8 „	L.S.	Hemsalech.
...	...	4426·5 „	4426·2 „	„	„
4421·5 „ (?)	...	4478·5 „	4478·0 „	„	„
...	...	4552·1 „	4552·3 „	„	„
...	...	4591·9 „	4590·0 „	„	Neovius.
4652·7 „	4651·1 „	4650·8 „	4651·2 „	N.B.	Hasselberg.
4707·7 „	4708·7 „	4708·3 „	4708·6 „	„	„
...	...	4779·2 „	4779·0 „	L.S.	Thalen.
...	...	4857·4 „	4860·6 „	„	Hemsalech.
...	5578·2 „	5578·4 „	?		
		5925, 6465 } bands.	Probably pos. bands of N.		

We see that the auroral spectrum is almost entirely due to nitrogen. The lines partly belong to the negative, partly to the positive band spectrum, and in the visible part also some lines appear, which in the literature are arranged in the line spectrum.

The grouping of these lines may be a more or less artificial one. In the northlight spectrum, which corresponds to a fairly definite way of production, they appear mixed together and with a quite typical intensity distribution. In the northlight spectrum they may be said to form one family of lines.

In all, 35 lines have been measured: out of these, 29 lines have been identified as belonging to nitrogen, and the two lines or bands 5925 and 6465 are probably N bands. Still, however, the following four lines—the green line included—are not yet interpreted:

5578·2
4182·5
3432·7
3208·3

These lines are not found among the recorded lines of the N spectrum. They cannot be ascribed to hydrogen, nor to helium or oxygen. I think there is little doubt that these lines also are due to nitrogen. This only means that the auroral spectrum is formed under conditions which are very difficult to reproduce in the laboratory. Several facts go to support this view, and I hope to gather some more knowledge through the continued work on the auroral spectrum.

A most important result of these observations of the auroral spectrum is that no indication of either hydrogen or helium lines has been observed. This fact is the more remarkable because in some of the spectra the stronger lines were greatly over-exposed.

During the last two years I have made experiments in my laboratory, the object of which was to study the light excited by the bombardment of cathode rays in mixtures of nitrogen and hydrogen, and nitrogen and helium. As a result of these investigations, which will be more fully treated later on, we can say that in mixtures N–H the presence of a few per cent. (3–7 per cent., say) of H can be detected in the spectrum when the N spectrum appears on the plate with about the same strength as that of the auroral spectrum. In mixtures of N and He the presence of 30 per cent. He can easily be detected.

It might be suggested that perhaps the energy possessed by the electric rays was sufficient for exciting the N spectrum, but too low to excite the H and He spectra: if so, these gases might be present in the higher strata of the atmosphere and still give no light.

But such an explanation is not possible. The energy of

the rays must be less than that corresponding to a potential fall of 30 volts; but whatever may be the nature of the cosmic rays, they must have a much greater energy if they get down to a height of 100 km. above the ground. Even if we only take into account the mass of nitrogen to be traversed, a cathode particle should have an energy corresponding to a fall through several thousands of volts, in order to get down to a height of 100 km., and any kind of positive rays would require a much higher voltage.

The absence of H and He lines in the auroral spectrum therefore shows that in that region where the auroral light is emitted, the pressure of H and He must be small compared with that of N in the same region.

Now, we know from observations of the height of aurora and the light distribution along the auroral rays* that the principal part of the auroral light is emitted in the height interval of 100–120 km. Taking the mean value as representing the region of emission, we must have that at the height of 110 km. the nitrogen pressure must be at least 15 times as great as that of hydrogen and 3 times as great as that of helium.

The possible quantities of H and He which may be present will depend on the assumption we make with regard to the pressure distribution of nitrogen. Now, the quantity of nitrogen present at the various heights will depend on the way in which the temperature varies upwards.

In previous papers dealing with the absorption of electric rays in the atmosphere†, I have, in accordance with Wegener, assumed that the temperature up to a height of 10 km. on an average can be put equal to $-23°$ C., and above this height put equal to $-53°$ C.

In Table IV. is given the pressure at various heights (h) for the gases H, He, N, and O, corresponding to the assumptions of Wegener.

At the height of 110 km. the pressure of nitrogen should be 0·055 dyn./cm.^{2}, and for H and He 10·25 and 0·494 respectively. But if this value for the N pressure were nearly right, the auroral spectral analysis would show that the pressure of H and He could not be greater than 0·0037 and 0·015 respectively, or the hydrogen pressure could only be about 4/10,000, and the He pressure only a 4/1000 part of that assumed by Wegener.

* L. Vegard and O. Krogness, "The Position in Space of the Aurora Polaris," Geophys. Publ. vol. i. No. 1, p. 149.

† Phil. Mag. vol. xlii. p. 47 (1921).

TABLE IV.

h. T=220.	*h'*. T=300.	Pressure in dyn./cm.^{2}			
		Hydrogen.	Helium.	Nitrogen.	Oxygen.
400×10^5 cm.	$542{\cdot}0\times10^5$ cm.	0·44	0·001	$4{\cdot}4\times10^{-21}$	...
300 „	405·6 „	1·30	0·008	$1{\cdot}69\times10^{-14}$	...
200 „	269·2 „	3·85	0·071	$6{\cdot}5\times10^{-8}$	...
160 „	214·5 „	5·95	0·169	$2{\cdot}8\times10^{-5}$	...
140 „	187·3 „	7·29	0·259	0·00058	...
130 „	173·7 „	8·24	0·321	0·0026	...
120 „	160·0 „	9·20	0·399	0·0121	0·0003
110 „	146·4 „	10·25	0·494	0·055	0·0013
100 „	132·8 „	11·4	0·611	0·251	0·0080
95 „	125·9 „	12·1	0·681	0·520	0·019
90 „	119·1 „	12·7	0·757	1·14	0·045
85 „	112·3 „	13·5	0·843	2·43	0·107
80 „	105·5 „	14·2	0·939	5·21	0·254
75 „	98·7 „	15·0	1·045	11·1	0·602
70 „	91·8 „	15·9	1·16	23·7	1·44
65 „	85·0 „	16·8	1·3	50·5	3·42
60 „	78·2 „	17·6	1·45	107·7	8·10
55 „	71·4 „	18·7	1·60	229·9	19·3
50 „	64·6 „	19·7	1·79	491·5	45·8

When the values of the pressures of H and He are given at 110 km., we can find the pressure at other heights, provided that we know the temperature. In Table V. is given

TABLE V.

h. T=220.	Pressure in dyn./cm.^{2}		
	Nitrogen.	Hydrogen.	Helium.
400×10^5 cm.	$4{\cdot}4\times10^{-21}$	0·00016	0·00004
300 „	$1{\cdot}69\times10^{-14}$	0·00047	0·0003
200 „	$6{\cdot}5\times10^{-8}$	0·0014	0·0026
160 „	$2{\cdot}1\times10^{-5}$	0·0021	0·0063
140 „	0·00058	0·0026	0·0096
130 „	0·0026	0·0030	0·012
120 „	0·0121	0·0033	0·015
110 „	0·055	0·0037	0·018
100 „	0·251	0·0041	0·023

the variation of the H and He pressure above 100 km., provided that we put the pressure at 100 km., equal to 0·0037 and 0·018 for the two gases respectively. But we notice that going upwards the pressure of H and He will already, at a height of 120–130 km., overtake that of nitrogen. We

should get hydrogen and helium lines as soon as we got a bit further up. In consequence, we should expect to find a change in the appearance of the auroral rays when they are seen to pass from heights of, say, 3–400 km. down to about 110 km. But no such change is to be noticed. And it seems as if there should not be any possibility of the existence of a H and He layer in the upper strata of the atmosphere.

Now we may be aware of the possibility that our calculations of the nitrogen pressures are based on false assumptions with regard to the temperature distribution.

Recently a most interesting paper has been published by Lindemann and Dobson *, dealing with the meteors. They have tried to calculate the amount of gas which a meteor must traverse in order to obtain the temperature necessary for its behaviour at the various heights.

They come to the conclusion that the density at a height of, say, 100 km. must be considerably greater than that ordinarily found, and they conclude that the higher strata must have a temperature of about 300° absolute.

Now we can easily calculate the pressures at various heights when we know the composition of the atmosphere 10 km. above the ground, and suppose that above this height the temperature is 300° absolute.

The pressure is found from a formula of the form :

$$p_h = p_0 e^{-\frac{n}{T}(h-h_0)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Let h be 10 km., then p_0 is the pressure at 10 km. If, now, we have calculated the pressure for one temperature, we can easily find the pressure distribution at another temperature in the following way :—

At the temperature T and a height h the pressure will be equal to that at a certain height h' at the temperature T_1 provided that

$$\frac{h'-h_0}{T_1} = \frac{h-h_0}{T} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

We have, in other words, by means of this formula to calculate the height h' which, for a temperature T_1, corresponds to a height h and a temperature T.

The heights h' corresponding to 300° are given in the second column of Table VI.

* F. A. Lindemann and G. M. B. Dobson, "A Theory of Meteors and the Density and Temperature of the Outer Atmosphere to which it leads," Proc. Roy. Soc. A. cii. p. 411.

TABLE VI.

h. T=300.	Pressure in dyn./cm.^{2}		
	Nitrogen.	Hydrogen.	Helium.
$542{\cdot}0\times10^5$ cm.	$4{\cdot}4\times10^{-21}$	0·0053	0·00096
405·6 ,,	$1{\cdot}69\times10^{-11}$	0·016	0·0077
269·2 ,,	$6{\cdot}5\times10^{-8}$	0·046	0·068
214·5 ,,	$2{\cdot}8\times10^{-5}$	0·072	0·162
187·3 ,,	0·00058	0·089	0·248
173·7 ,,	0·0026	0·099	0·308
160·0 ,,	0·0121	0·111	0·383
146·4 ,,	0·055	0·123	0·475
132·8 ,,	0·251	0·138	0·587
125·9 ,,	0·520	0·145	0·654
119·1 ,,	1·14	0·153	0·727
112·3 ,,	2·43	0·162	0·810
105·5 ,,	5·21	0·170	0·902
98·7 ,,	11·1	0·180	1·004

If, now, in the same way as before, we apply the results from our auroral spectral analysis on the distribution of nitrogen, which corresponds to T=300°, we get the somewhat higher values for the possible pressures of H and He given in Table VI.

Also in this case, however, H and He would predominate above a height of 120–140 km.

As the ordinary hydrogen and helium lines are not to be found in the auroral spectrum, I think we can safely conclude that the green line (5578) cannot originate from any of these two gases, for this line remains the most prominent to the very bottom edge of all ordinary auroræ going down to a height of 95–100 km.

But at this height—as we saw—the pressure of H and He is very small as compared with that of nitrogen: and remembering that the N spectrum is very easily excited, it can hardly be assumed that a gas which is only present with a few per cent. in the mixture shall give the strongest line.

On the other hand, I have found by spectral observations that the green line is seen to the very top of the auroral ray streamers, and the gas which emits this line must be a prominent component of the atmosphere up to its extreme upper limits.

Hence we conclude that the hydrogen and helium layer, which has earlier been supposed to dominate at the top of our atmosphere, does not exist.

For the same reason we can hardly assume the existence

of some light unknown gas (geocoronium); for this gas had to be present at a height of 100 km., with a pressure of the same order of magnitude as that of nitrogen, in order to give the most prominent of all lines in the auroral spectrum. Being a light gas, however, geocoronium should soon predominate, and we should expect that any trace of a N spectrum should disappear at a height of, say, 140 km.

This conclusion is independent of the assumption we make about the temperature of the upper strata of the atmosphere, for if the temperature is high the pressure of N at a height of 100 km. will be greater, and for geocoronium we had to assume a correspondingly great pressure to make its green line predominant at this altitude. Somewhat higher up in the atmosphere the lighter gas (geocoronium) would take the lead.

If, then, we were able to show that in the auroral spectrum the nitrogen lines are maintained at a height greater than 150 km., I think we were justified in concluding that also the green line (5578) must be due to nitrogen. Observations are now in progress for testing this point.

From a physical point of view, however, it is very unlikely that a new gas (geocoronium) should exist, because there is no place for it in the periodic system.

I therefore think the best procedure would be to try to interpret the results of our spectral analysis without introducing the possibility of a new type of matter. The introduction of geocoronium is merely an easy way of getting out of the difficulty.

But if we give up geocoronium, the green line must be ascribed to nitrogen, and nitrogen must be the predominant gas up to the very limit of the atmosphere.

Here, however, we meet with a difficulty of another nature. From measurements of the height and position of the auroræ, we know that sometimes the upper end of an auroral ray can reach a height of 500–600 km., and at this height its light intensity is of the same order of magnitude as further down towards its bottom edge.

In previous publications * I have dealt with the variation of light intensity along the ray streamers. The increase of intensity upwards can be explained by assuming that a greater part of the cosmic rays when they enter the atmosphere form great angles with the magnetic line of force. In this way we may explain that intensity variations along

* L. Vegard and O. Krogness, 'The Position and Space of the Aurora Polaris,' p. 149. L. Vegard, Phil. Mag. xlii. p. 59.

the streamers may occur, but in order to get a noticeable intensity it is of course necessary that the pressure of the atmosphere does not fall below a certain limit.

From Table IV. we see that the nitrogen pressure decreases rapidly upwards. Assuming a temperature of 220° absolute above 10 km., the nitrogen pressure at a height of 400 km. should have the extremely small value of $4{\cdot}4 \times 10^{-21}$ dyn./cm.^{2}

In this respect it does not help much to assume a somewhat higher temperature. With a temperature of 300° observed, the same low pressure would exist at a height of 542 km., which is also inside the auroral region.

If we at all are allowed to speak of a pressure of this order of magnitude, it would mean that at a height of 400–600 km., where the pressure is of the order 10^{-21} dyn./cm.^{2}, there should only be one molecule in a volume of 10 cubic metres. This, again, would mean that the density of the electric radiation had to be enormously great to produce the light intensity actually observed.

If, *e. g.*, we assume that the auroræ are produced by cathode rays of a certain velocity, $v = \beta c$, we can estimate the density of radiation by a comparison with the light produced when nitrogen is bombarded by a stream of cathode rays in a vacuum tube.

We suppose the sources of light to be placed before the slit of a spectroscope at distances R and r from the slit. The effective cone of the instrument with a solid angle ω cuts the sources in areas S and s. Let the intensity of light per unit area of the northlight and the vacuum tube be J_a and J_v. The quantity of light passing through the instrument in unit time will be proportional to

$$\frac{qSJ_a}{R^2} \quad \text{and} \quad \frac{qsJ_v}{r^2}$$

for the auroræ and the vacuum tube respectively: q is the area of the slit, and

$$\frac{S}{R^2} = \frac{s}{r^2} = \omega.$$

If the two sources produce the same effect in the instrument, we get

$$J_a = J_v. \qquad \qquad (3\,a)$$

Now, the intensity per unit area is equal to the intensity per unit volume multiplied by the thickness of the layer, because there is practically no absorption in the layer, at any

rate in the visible part of the spectrum. The condition of equal photographic effect then will be

$$i_a l_a = i_v l_v \; ; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3\,b)$$

i is the intensity per unit volume, l is the thickness of the layer. The quantity i is proportional to the number of collisions made by the cathode particles per unit volume in unit time or proportional to νp, where ν is the number of electrons crossing unit area in unit time. Let n be the number of cathode particles which at any moment is present in 1 cm.^{3}; then

$$n = \frac{\nu}{v}.$$

Assuming the same ray velocity, the intensity of light emission per unit volume will be proportional to np, and our condition for equal spectroscopic action take the form

$$n_a p_a l_a = n_v p_v l_v. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

For the vacuum tube we assume $l_v = 1$ cm., $p_v = 100$ dyn./cm.^{2} The corpuscular current which would produce the same spectroscopic action as that of the upper part of an auroral ray can be estimated to about 10^{-6} ampere when the velocity of the rays is about $\frac{1}{2}c$, where c is the velocity of light.

This would give

$$n_v e \beta = 10^{-7}.$$

Putting $\beta = \frac{1}{2}$ and $e = 4{\cdot}8 \times 10^{-10}$,

$$n_v = 4 \times 10^2,$$

Hence $$n_a p_a l_a = 4 \times 10^4;$$

l_a is of the order of 10^6 cm., and

$$n_a p_a = 4 \times 10^{-2}.$$

If, now, p_a were equal to 4×10^{-21}, it would follow that

$$n_a = 10^{19}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

At any moment there should be about as many electrons (ray carriers) present in each cm.^{3} as there are molecules in 1 cm.^{3} of gas at 0° C. and atmospheric pressure. Such a density of electric radiation cannot be assumed. The electrostatic forces would prevent a ray bundle of such a density being formed.

If nitrogen is the predominant gas to the very limit of the atmosphere, it must at the height of 400–600 km. possess a density of a much higher order of magnitude than that previously calculated.

Quite formally we could get a sufficient pressure and density of nitrogen at a height of 600 km. by a proper assumption with regard to the temperature of the upper strata of the atmosphere.

If, as before, for the sake of simplicity, we suppose the temperature to be constant above 10 km. from the ground, we can easily find the temperature which is necessary to make the nitrogen pressure at 600 km. equal to a given pressure.

From Table IV. we see that for a temperature of 220° absolute, the pressure of nitrogen already at a height of 130 km. is only 0·0026 dyn./cm.^{2} If a nitrogen pressure of this magnitude should exist at a height of 600 km., the corresponding temperature T can be found from equation (2) by putting $h=130$, $h_0=10$, $T=220°$, $h'=600$; and we get

$$T=1078° \text{ abs.}$$

That the atmosphere above 10 km. should have a temperature of this magnitude must, I think, be considered as excluded.

The simplest way in which to prevent the nitrogen density from diminishing so rapidly as we pass upwards in the auroral region, would be to suppose that the upper strata were electrically charged, and consequently were acted on by electric forces. We might suppose the gas near the limit of the atmosphere partly to exist as positive ions. This is, in fact, what we should expect from a physical point of view. The upper layers of the atmosphere are exposed to the direct action of the sun's radiation. On account of the photoelectric effect, electrons will be driven out from the gas molecules with maximum velocities determined by the Einstein equation

$$\tfrac{1}{2}mv^2=h\nu, \quad \ldots \ldots \quad (6)$$

where ν is the maximum frequency of the incident light, and h is Planck's constant.

Now it is pretty certain that the sun—besides the ordinary light spectrum—emits radiation of much shorter wave-length of the type we know from the X- and γ-rays. With energy quanta of this magnitude, electrons may be driven out of the atmosphere from quite a considerable layer of gas round the earth, leaving the gas molecules behind in the form of positive ions.

In this way a positively charged shell will be kept round the earth. Above a certain height the electric force will be directed upwards, and below this height it must, on account of the negative charge of the earth, be directed downwards.

The variation of pressure will no longer be given by the equation

$$dp = -\rho g\, dh, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

but instead of this we now get for higher strata

$$dp = -(\rho g - \sigma \mathrm{F}) dh, \quad . \quad . \quad . \quad . \quad , \quad (8)$$

where σ is the electric charge per unit volume and F the electric force. The electric charge will diminish the weight of a given quantity of gas; it will, as it were, make the molecules lighter. In this way we can understand that nitrogen in the auroral region can be distributed as if it were one of the very lightest gases.

Light gases like hydrogen and helium when they get ionized may have their weight so much reduced that they would fly away from the earth; and thus we may explain the absence of hydrogen and helium layers which have earlier been assumed to exist on the top of our atmosphere.

If the upper strata of the atmosphere to a great extent consist of positively-charged N molecules, we can also understand that under this condition the bombardment with electric rays may produce other N lines that we observe in our laboratory experiments, for it only means that we have not been able to produce artificially in the observation chamber a quantity of nitrogen containing a sufficiently large percentage of positive ions.

I also think that this hypothesis of an electrically-charged upper layer gives a very simple explanation to the results found by Lindemann and Dobson from their investigations of the meteors, for the greater density which they want will be produced by the electric charge of the upper layers, and we need not take refuge in the very improbable hypothesis of the high temperature which they suppose to exist above 10 km.

The highly ionized state of the atmosphere must be restricted to a certain layer, and as long as the cosmic rays are absorbed in this layer, we get the ordinary auroral spectrum showing the typical green line. As the typical greenish-yellow auroræ may have their bottom edges so far down as 100 km. above the ground, this would mean that the positive layer should go down to a height which is not greater than 100 km.

If the penetrative power of the rays could be great enough to enable them to pass into the neutral atmosphere, the nitrogen spectrum would turn into the more ordinary type and the green auroral line should disappear.

As a change of spectrum must be accompanied with a change of colour, this hypothesis of an electrically-charged upper layer gives us new possibilities for explaining the marvellous changes of colour which the auroræ may display.

Thus I think that the peculiar colour of a drapery-shaped arc, which I observed at Bossekop, October 1912, may be simply explained in this way. From the upper limit to a distance of a few km. from the bottom edge the arc had the ordinary greenish-yellow colour, but at a certain height the colour turned into dark red. This red bottom edge was found simultaneously all along the arc, which extended across the sky from E.N.E. to W.N.W., and was observed at least 8 minutes.

It would be of great interest to measure the height of such forms to see if they come lower down than the ordinary green arcs; but these phenomena are very rare, and the height of such arcs has not yet been measured. If our view is right, however, such auroræ would show us where the neutral atmosphere begins.

The auroræ, however, also show colour changes of another type. Separate auroral streamers or ray bundles forming part of a drapery or drapery-shaped arc may turn into another colour, usually red or bluish-red. The colouring is in this case not restricted to the bottom edge and is not kept for a considerable time, but it extends to almost any part of the streamer, and sometimes the whole ray bundle may turn red.

The red streamers appear in between the ordinary green-yellow ones; usually they only last for a fraction of a second. Other red-coloured bundles shoot in, and the distribution of colours, and thus the whole picture, will undergo very rapid changes.

To explain this kind of colour variation, I have in previous papers introduced the hypothesis that usually only a fraction of the rays that enter the atmosphere get absorbed, so that most of the rays return into space. Remembering that the spectrum from a gas bombarded by cathode rays depends on the velocity of these rays, we may expect to get a variation of colour according as a greater or smaller fraction of the rays are completely absorbed.

Some of the colour changes may no doubt be caused in this way; but if the upper layer of the atmosphere is positively charged, there will also be another possibility for colour changes to take into consideration.

If a very intense ray bundle (consisting of negative electrons) pass through the positively-charged layer, and if it keep its position for some short time, the gas on its way may for a moment become partly neutralized, and the N spectrum may partly turn into the ordinary type corresponding to neutral nitrogen and which gives the red colour.

It is of interest to notice that the latter explanation of colour changes would require negatively-charged cosmic rays. In fact, the existence of a positively-charged upper layer of gas would involve that, at any rate, all auroræ showing a ray structure must be produced by rays of negative electrons. For, as we saw, the electrical charge of the layer would increase the density of matter throughout the auroral region. The mass of gas in a cylinder of unit cross-section reaching from a height H to infinity will be

$$m_{\mathrm{H}}=\int_{\mathrm{H}}^{\infty} \rho\, dh. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

For a neutral gas we have $m_{\mathrm{H}}=\frac{p_{\mathrm{H}}}{g}$; but this relation does not hold for an electrically-charged layer. Integrating the equation (8), we get

$$-\int_{\mathrm{H}}^{\infty} dp=g\int_{\mathrm{H}}^{\infty} \rho\, dh-\int_{\mathrm{H}}^{\infty} \mathrm{F}\sigma\, dh,$$

$$m_{\mathrm{H}}=\frac{p_{\mathrm{H}}}{g}+\frac{1}{g}\int_{\mathrm{H}}^{\infty} \mathrm{F}\sigma\, dh. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

This equation shows that the mass m_{H} which a ray has to penetrate to get down to a height H is now greater than that calculated for neutral nitrogen (Table IV.); and if the charge is to produce a sufficient density of nitrogen 600 km. above the ground, the mass m_{H} at the height of 100 km., say, must be very much greater than that found for neutral nitrogen.

Any electric ray with a carrier of atomic dimensions which had a penetrating power great enough to enable the ray to get down to a height of 100 km., would possess a magnetic deflectibility, which would be too small to explain the narrow ray streamers.

Only electron rays combine a sufficient penetrating power with a sufficient magnetic deflectibility to explain the height and structure of the aurora *.

Physical Institute, University,
Christiania,
March, 1st, 1923.

XX. *The Theory of the Abnormal Cathode Fall.*

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

I HAVE read with great interest a paper on the above subject published in this journal last month by the Research Staff of the General Electric Company whom I shall, for the convenience of this letter, identify with Mr. Ryde as his name is associated with the work. In consideration of the fact that the theory put forward is based almost entirely upon measurements made by me, I feel justified in making a few comments to caution the reader that its novelty of treatment and agreement with experiment are more apparent than real.

Under the more usual title of "Dark Space" I studied this fascinating phenomenon continuously for ten years, and I am entirely unable to accept Mr. Ryde's view that this simple theory is adequate even for a preliminary survey of the facts. To mention two minor points at the outset, I cannot see how a theory depending solely on the mass of the positive ions can give two entirely different results in the cases of N_2 and CO since we expect, and positive ray analysis proves, that the mean mass of the positive ions is practically identical in both. Again, with the definite stated conditions under which all my measurements were made, I never found the slightest evidence of a measurable potential

* *Note added to the Proof.*—Since I wrote this paper observations have been made with regard to the auroral spectra emitted at various altitudes, and investigations have been continued regarding the constitution of the upper strata of the atmosphere.

These investigations have led to the view that a highly ionized upper layer cannot exist in the form of gas—but we must assume the charge mainly attached to clusters—or small crystals of nitrogen. These investigations will be dealt with in a second paper.

At present I should only like to point out that in an upper layer formed by dust particles more or less electrically charged we can no longer apply the gas equations (1), (7), and (8). The conclusions based on the assumption that these equations hold for the auroral region therefore ought to be reconsidered.

difference between the metal of the anode and the edge of the negative glow. If such exists, direct experiments show that it cannot be more than a few volts, which will not go far to explain the discrepancy of about 130 volts shown by Mr. Ryde's parallel curves.

In the first paper I published on the subject in 1907 I gave a simple theory, which Mr. Ryde appears to have overlooked, founded on precisely the same premises as his, namely, that the boundary of the negative glow could be regarded as a highly conducting plane source of positive ions and that the density of electrons in the dark space could be neglected in comparison with that of the positive ions. The only difference between my theory and the one put forward is that I took for the velocity of the positive ions the expression for their mobility at ordinary pressures, whereas Mr. Ryde takes that of the free fall *in vacuo*. When I formulated my theory I had the choice of either of these expressions, but considered the latter certain to give too high a result. In this choice I have since been justified, for the difficulty is in making the velocity *low* enough to fit the facts. My theory required id^3V^{-2} to vary inversely as the pressure, which agreed very well with the numerical results obtained, but I abandoned it without any hesitation on ascertaining the law of distribution of potential in the dark space in 1911. The method by which the latter very simple but utterly baffling result was then demonstrated appears entirely free from objection. I have no more reason to doubt its substantial accuracy now than I had when I made the measurements, so that I find myself unable to consider very seriously any theory which does not attempt its explanation.

Equally formidable obstacles to the formulation of a workable theory are raised in my paper on the effect of the material of the cathode, where it is shown that a silver cathode gives twice the length of dark space given by a magnesium one and that with the same current and pressure the relation between V and d is accurately linear for six out of the nine metals tried. Further experiments with perforated cathodes carried out before the war but not published till 1919, showed that the phenomenon just in front of the cathode, far from being a simple hail of positively charged particles, was inextricably complicated by intense local ionization. I was therefore forced to the conclusion, which has been further strengthened by my subsequent work on positive rays, that no simple theory can explain the mechanism of the discharge in a satisfactory quantitative manner. It

seems in this way as intractible as an ordinary gas-flame, to which it bears many striking resemblances.

In conclusion, I may state that I now think that the true solution of all these perplexing discrepancies is probably to be sought for in the fact, now well established, that positive ions moving rapidly through a gas do not retain their identity but gain and lose electrons t e whole time as they collide with other particles. The mechanism of these exchanges and its effect on the current carried is entirely unknown, but experiments of several different kinds are now focussed on this problem which should yield very valuable results in the near future.

F. W. Aston.

June 7th, 1923.

XXI. *On the Motion of Electrons in Gases. By* V. A. Bailey, *M.A., D.Phil., Demonstrator, The Electrical Laboratory, Oxford, Lecturer of Queen's College, Oxford* *.

1. A LARGE number of investigations have been made of the motion of electrons in gases, and it seems desirable to point out what reliable information may be obtained from the experiments, as the present generally accepted views of the behaviour of electrons in gases at the higher pressures and under low forces are to a great extent unsatisfactory. Too frequently the results of different experimenters do not agree with one another, or are obtained by a wrong interpretation of the observations; while in many cases the methods used are fundamentally unsound.

For values of the electric force X and gas-pressure p less than those for which ionization by collision occurs in notable quantities, the measurements are mainly concerned with W the velocity in the direction of the electric force, and K the coefficient of diffusion of charged particles in a gas. The particles may in general be either electrons or ions, but the motion of electrons is of special interest. The conditions under which ions are formed by electrons adhering to molecules have also been the subject of much discussion.

In some cases with ions the velocity W is found to be proportional to the ratio X/p. The constant of proportionality may then be termed the mobility of the ions, although the more common practice is to call W/X by that name, the gas being at atmospheric pressure.

* Communicated by Prof. J. S. Townsend, F.R.S.

No useful object is served by extending this definition to those cases where W is not proportional to X/p.

Some experimenters have made the mistake of basing their experiments on the assumption that W is always proportional to X/p, without verification. Considerable errors have arisen from this cause, for in the case of electrons the velocity W is generally not proportional to the ratio X/p.

2. Methods * for measuring W have been devised by Rutherford, Langevin, Zeleny, Chattock, and Townsend. The methods of Zeleny and Chattock are restricted to those cases where W is proportional to X, and the same is true of that particular type of Rutherford's method in which an alternating e.m.f. is used. But in Lattey's form of Rutherford's method and in Langevin's method no such restriction occurs, and reliable results may be obtained by them, except in those cases where diffusion effects are comparable with those due to the motion W. The effects of diffusion may be avoided by Townsend's method which is particularly suitable for determining the values of W for electrons.

The only methods of determining the coefficient of diffusion K are those due to Townsend, which are described in his book on 'Electricity in Gases,' Chapter V.

3. When in 1908 † it was shown for the first time that in dry gases (or even in moist gases, provided X/p is sufficiently high) the ions exist as free electrons having an abnormal energy of agitation as a result of elastic collisions with the gas molecules, it became at once an important matter to determine the velocities W corresponding to this novel state of affairs. This was done by Lattey ‡ for air. and by Franck § for argon and nitrogen, both using modifications of Rutherford's method.

In Lattey's form the charged particles coming through a gauze move under a constant electric force towards a parallel gauze for a known short interval of time t, and are then swept back again by reversing the field. No charged particles reach the second gauze until t is increased up to the value L/W, where L is the distance between the gauzes.

Franck made use of a sinusoidal alternating field in place of Lattey's constant field, so that his particles traversed the distance L with a variable velocity. He was thus compelled to assume that the velocity W was always proportional to X,

* *Cf.* J. S. Townsend's 'Electricity in Gases,' Chapter IV.

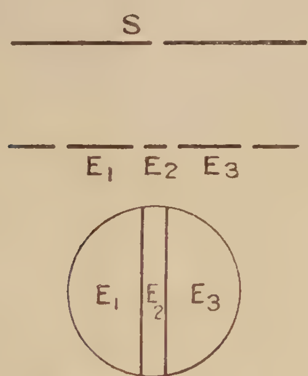
† J. S. Townsend, Proc. Roy. Soc. A. lxxx. p. 207, and A. lxxxi. p. 464 (1908).

‡ R. T. Lattey, Proc. Roy. Soc. A. lxxxiv. p. 173 (1910).

§ J. Franck, *Verh. d. Deut. Phys. Ges.* xii. pp. 291, 613 (1910).

which is shown very emphatically not to be generally true of electrons by Lattey's experiments and by the more recent experiments * with Townsend's method. We shall see later that another disadvantage of the alternating field method is that it gives misleading information about the effect of impurities like oxygen which may form ions from free electrons.

There is a further source of error to which the Rutherford method is liable when used for studying the motion of free electrons, even when the constant field modification is adopted. A small group of electrons starting from a point near the first gauze do not all arrive simultaneously at the second gauze, but, owing to the effect of diffusion, the group spreads out in all directions as it moves, and electrons arrive at the gauze before the group as a whole traverses the distance L. Thus the observed time t will depend somewhat on the sensitiveness of the instruments used for detecting the first signs of electricity arriving at the second gauze, and also on the magnitude of the diffusion when compared with the velocity in the direction of the field.



Townsend's method of measuring W is, briefly, as follows: the ions or electrons (moving in a uniform field X throughout) pass through a narrow slit S, shown in the figure, and arrive on the electrodes E_1, E_2, E_3, which are separated by two narrow air-gaps. The centre of the stream falls on the centre of E_2. A suitable uniform magnetic field H is applied

* J. S. Townsend and H. T. Tizard, Proc. Roy. Soc. A. lxxxviii. p. 336 (1913). J. S. Townsend and V. A. Bailey, Phil. Mag. xlii. Dec. 1921; xliv. Nov. 1922. M. F. Skinker, Phil. Mag. xliv. Nov. 1922.

perpendicular to X and parallel to the slit (or electrode E_2), thus deflecting the centre of the stream towards one of the air gaps. The angle of deflexion θ is known, and W can then be calculated by means of the simple formula $HW/X = \tan\theta$.

This method is free from the defects inherent to Rutherford's method which have just been considered, and is particularly well adapted to the case of free electrons, since the high values of W for these latter allow the use of moderate magnetic fields.

4. The evidence of the possible existence of free electrons in a gas was obtained by Townsend * in the course of determinations of the quantity *Ne* † for different gases. He showed that in general this free existence depends on three main factors: the nature of the gas, its state of dryness, and the value of X/p. The larger the percentage of water-vapour present, the larger is the minimum value of X/p necessary to keep electrons in the free state, and even in pure water-vapour it is possible to maintain this condition of the electrons.

The experiments of Townsend and Tizard ‡ established that in very dry air the free electronic state exists when X/p exceeds the value 0·2, and the more recent experiments on very dry N_2, H_2, A, He §, and CO_2 ‖ give similar results.

The case of O_2 requires special mention, as there appears to be a general impression that small traces of this gas in any other gas will prevent the existence of free electrons. This conclusion appears to have been arrived at for the first time by Franck, and in support of it he gives the results of his mobility measurements for argon and nitrogen—namely, that the addition of 1·2 per cent. or more of oxygen to pure argon reduces the mobility of the negative ions from 206 to 1·7, and that a similar addition of oxygen to nitrogen reduces the mobility from 145 to 1·84. These conclusions as they stand are in agreement neither with Lattey's experiments with air nor with those which were subsequently made by Townsend's method with air and oxygen. It appears from the investigations with oxygen that the electrons tend to combine with the molecules of that gas only at small values of X/p; and the presence of the normal 20 per cent. of

* J. S. Townsend, *loc. cit.* (1908).

† N = number of molecules in 1 c.c. of gas at 15° C. and 760 mm. Hg; *e* = charge on an ion.

‡ *Loc. cit.*

§ J. S. Townsend and V. A. Bailey, *loc. cit.*

‖ M. F. Skinker, *loc. cit.*

oxygen in the air used by Townsend and Tizard did not appreciably affect the freedom of the electrons when X/p exceeded 2.

The erroneous conclusions of Franck may be attributed largely to the fact that the form of the Rutherford method which he used is very misleading for the purpose of investigating this point. The electrons in his apparatus moved in an alternating electric field, and were consequently in very weak fields for part of the time. It would be during this interval of time that the electrons became attached to the oxygen molecules, and as this would occur more or less at all values of the field amplitude, it would appear from these experiments that the oxygen molecule attaches electrons to itself under all conditions.

It may then be concluded that Franck's form of the Rutherford method is in general quite unsuitable for the study of the motion of electrons. Lattey's form of the method, however, is much less open to criticism, and fails to give accurate results only when the effects of diffusion become appreciable; but the method was originally designed for measuring the velocities of ions, and cannot be easily adapted to measure the higher velocities of electrons.

5. A rough general outline of the present state of knowledge about the motion of electrons in gases with low values of X/p, may be presented as follows. In order to obtain reliable values of the mean free path of electrons and the energy lost in collisions, it is necessary to determine experimentally both the velocity W in the direction of the electric force and the velocity of agitation u of the electrons. Theory and experiment agree in showing that when the electric force X and the gas-pressure p are varied so as to keep the ratio X/p constant, then the velocity W and the kinetic energy of agitation of the electrons remain constant, *i.e.* $W=f(X/p)$ and $k=\phi(X/p)$, where k is the factor by which the kinetic energy of the electrons exceeds the kinetic energy of an equal number of molecules of gas at a standard temperature. The functions f and ϕ are characteristic for each gas, but possess no simple mathematical form.

Regarding the equations $W=f(X/p)$, $k=\phi(X/p)$ as established experimentally, it is possible by means of them to find the mean free path l of the electrons in the gas (at some standard pressure), and the mean energy E lost by an electron at each collision with a gas molecule. Both l and E are functions solely of the mean velocity of agitation u of the electrons.

In those gases where the electrons can attach themselves

to the molecules under certain conditions, we may in the same manner consider that the probability of an electron adhering to a molecule is a function of u only. If in N collisions made by the electrons hN result in attachment, we may call h the probability of ion formation and may set $h = F(u)$ where F is characteristic for each gas. Very little is known about h, except that it is practically zero for N_2, H_2, He, A, and CO_2. Electrons may travel several centimetres through those gases, making several thousand collisions with molecules, and not become attached to them. In oxygen, water-vapour, and some other gases and vapours, h is not zero, except possibly for large values of u. In general, as appears from the experiments on ionization by collision, h is extremely small for large values of u.

This point of view differs from that of J. J. Thomson *, who considers h to be constant for a given gas, and thus independent of u. It is impossible to reconcile Thomson's view with the above experimental facts about oxygen and water-vapour.

L. B. Loeb † has attempted to determine h for a number of gases on the basis of Thomson's theory and using Franck's method of measuring the "mobilities." But as Thomson's theory conflicts with a large amount of experimental evidence, and Franck's method of measuring "mobilities" is very unsuitable for dealing with a stream containing *any* electrons, Loeb's results cannot be regarded as being convincing.

XXII. *Notices respecting New Books.*

Physiology of the Ascent of Sap. By Sir JAGADIS CHUNDER BOSE, M.A., D.Sc., F.R.S., C.S.I., C.I.E. Longmans, Green & Co. 1923. 16*s*. net.

IN this book Sir Jagadis Bose has collected together a series of the researches which he has been carrying out at the Bose Research Institution recently with the co-operation of a large number of research students. The problem of the ascent of sap has for a long time baffled researchers. The main difficulty of the problem resides in the lack of adequate means of detecting and measuring the rate of ascent of transpiration, exudation, and their induced variations. The elaborate and intricate instruments which

* J. J. Thomson, Phil. Mag. i. p. 369, April 1901.

† L. B. Loeb, Phys. Rev. xvii. 2 (1921), and Phil. Mag. xliii. Jan. 1922.

have been devised by Sir Jagadis Bose have been further developed for the attack on this problem. Various types of automatic recorders of great sensitiveness and precision have proved of value: the electric probe has been used to localize phenomena occurring in the cells.

XXIII. *Proceedings of Learned Societies.*

GEOLOGICAL SOCIETY.

[Continued from vol. xlv. p. 799.]

December 20th, 1922.—Prof. A. C. Seward, Sc.D., F.R.S., President, in the Chair.

THE following communications were read:—

1. 'A Micrometric Study of the St. Austell Granite (Cornwall). By William Alfred Richardson, M.Sc., F.G.S.

The problem of the effect of sampling a coarse-grained rock by means of slices is considered in detail, and to estimate this effect a statistical standard and the consistency of the mapped values are used.

Qualitative and quantitative study of the minerals reveals three types of rock: (*a*) a biotite-muscovite-granite of coarse grain confined to the east; (*b*) a lithionite-granite occupying by far the greater part of the outcrop; and (*c*) a gilbertite-granite confined to a small area near St. Stephen's Beacon, and furnishing the 'china-stone' rock.

The correlations of certain minerals are examined. A high negative correlation is found between quartz and orthoclase—true for this area, but not for granites in general. Almost complete correlation is indicated between topaz and minerals of the contact-group, while the correlation between biotite and tourmaline is hardly of any significance.

When mapped, the minerals fall into groups that show little relation to the boundaries of the granite as a whole, but are distinctly connected with the areas occupied by the different types. The minerals within the different areas are arranged in a similar way. There is an outer zone rich in quartz and to some extent in mica, surrounding an inner region noteworthy for a high content of orthoclase; while the plagioclase is sometimes concentrated centrally, and sometimes towards the margin.

There is evidence to show that the magma invaded the area progressively from the east to the west; and that it had always partly, and sometimes largely, crystallized before injection into the

present level. An explanation of the exceptional arrangement of the minerals, based on the motion of the invading magma, is suggested.

2. 'The Petrography and Correlation of the Igneous Rocks of the Torquay Promontory.' By William George St. John Shannon, M.Sc., F.G.S.

A description of the field relations is given, and it is demonstrated that two stages of vulcanicity occurred—in the Middle and in the Upper Devonian, as shown by basic tuffs and a spilite.

The intrusions form an alkaline suite. Albite-dolerite, with segregations, forms a laccolite at Black Head, and carries quartz and occasional olivine. Two isolated outcrops are correlated with this on petrographic grounds. Metasomatic silicification is ascribed to this intrusion. Evidence of the stability of the albitization and the subsequent alteration of augite to limonite is given: at the Red Rocks, Babbacombe, this approaches a laterite in character.

An augite-lamprophyre in limestone, and a soda-porphyrite in Middle Devonian slates are described from Babbacombe.

A preliminary account of the tectonics is attempted, particularly of the inversion, at Ilsham, of the faulting and of the north-to-south strike of some of the folds.

The results may be summarized as the recognition of the extension of the South-Western Alkaline Province; the existence of an alkaline intrusion, in addition to segregations; the Upper Devonian age of certain slates; and finally, the establishment of the apparent tectonic succession and post-Culm date of the folding.

January 10th, 1923.—Prof. E. J. Garwood, Sc.D., F.R.S., Vice-President; and afterwards Prof. A. C. Seward, Sc.D., F.R.S., President, in the Chair.

Prof. William Johnson Sollas, Sc.D., F.R.S., F.G.S., delivered a lecture on Man and the Ice-Age.

He said that, thanks to the researches of General de Lamothe, Prof. Depéret, and Dr. Gignoux, the Quaternary System now takes its place as a marine formation in the stratified series.

Four ancient coast-lines, of remarkably constant height, have been traced around the Mediterranean Sea and along the western shores of the North Atlantic Ocean. These, with their associated sedimentary deposits, form the successive stages of the Quaternary System: namely, the Sicilian (coast-line about 100 metres); the Milazzian (coast-line about 60 m.); the Tyrrhenian (coast-line about 30 m.); and the Monastirian (coast-line about 20 m.).

The Sicilian deposits rest unconformably upon the Calabrian

(Upper Pliocene), and in their lower layers contain a characteristic cold fauna. The fauna of the Milazzian is warm-temperate, of the Tyrrhenian and Monastirian still warmer, for they contain numerous species of mollusca which now live off the coast of Senegal and the Canary Islands.

The three lower coast-lines correspond with the three lower river-terraces of the Isser (Algeria), the Rhône, and the Somme. Hence it may be inferred that the position of the river-terraces has been determined by the height of the sea-level.

The lower gravels of the three lower terraces of the Somme all contain a warm fauna, *Elephas antiquus* and *Hippopotamus*, and thus (like the corresponding marine sediments) testify to a warm climate. The climate of the Quaternary age was, on the whole, warm-temperate or genial, but interrupted by comparatively short glacial intervals.

The outermost moraine (Mindel) of the Rhône Glacier is associated with the Milazzian terrace, the intermediate moraine with the Tyrrhenian, and the innermost moraine (Würm) with the Monastirian: except for their serial order, these associations are (in a sense) accidental.

It is now possible to assign the Palæolithic stages of human industry to their place in the Quaternary System: thus the 'Strepyan' or pre-Chellean is Milazzian in age; the typical Chellean—Tyrrhenian; the evolved Chellean, Acheulean, and Lower Mousterian—early Monastirian; and the Upper Mousterian, Aurignacian, Solutrian, and Magdalenian—later Monastirian.

The coast-lines of the Northern Hemisphere appear to have their counterparts in the Southern Hemisphere, and the researches of Dr. T. O. Bosworth in Peru and Prof. G. A. F. Molengraaff in the East Indies have revealed extensive marine Quaternary deposits and successive movements of the sea-level.

The Quaternary movements are probably due to a general deformation of the globe, involving eustatic changes in the level of the sea.

February 7th, 1923.—Prof. A. C. Seward, Sc.D., F.R.S., President, in the Chair.

Mr. G. Vibert Douglas delivered a lecture on the Geological Results of the Shackleton-Rowett (*Quest*) Expedition. The Lecturer said that St. Vincent and St. Paul's Rocks were examined on the way out, but the more detailed work commenced in South Georgia. This island lies 900 miles east of Cape Horn and is 100 miles long by 20 miles in width. Its topographical features are those of an upland dissected by glacial action. The glaciers in general show signs of withdrawal. Geologically,

the island is composed of sedimentary rocks and, at the south-eastern end, igneous rocks. These have been classified by Mr. G. W. Tyrrell as follows :

Sedimentary Rocks...	(1) Mudstones, shale, slate, phyllite.
	(2) Quartzite, greywacké.
	(3) Calcareous rocks.
	(4) Tufaceous rocks.
Igneous Rocks	(1) Gabbros and peridotite.
	(2) Dioritic and granitic rocks.
	(3) Dolerites and basalts.
	(4) Spilitic lavas and epidosites.

The question as to whether the sediments represent one continuous period of deposition is open to dispute. The Lecturer thought that there were two distinct periods, divided by an unconformity. Definite fossil evidence is difficult to obtain, but *Araucarioxylon* has been identified by Prof. W. T. Gordon, which would point to an age not older than Lower Carboniferous. This fossil came from the Bay of Isles, and was found in what the Lecturer believes to be the younger series. The rocks all show signs of metamorphism, and the strike of the folds and lamellæ of the phyllites would point to the fact that the pressure came either from the south-south-west or from the north-north-east. Considerable faulting was observed, both normal and reversed.

The igneous complex east of Cooper Bay can be differentiated into two separate areas : (1) north of Drygalski Fjord, and (2) at Larsen Harbour. In the former area quartz-diorite, peridotite, aplite, and syenitic lamprophyre with basement gabbro occur ; in the latter area were found spilitic lavas (containing much epidote) and basement gabbro. The general types are not Andean.

Elephant Island is situated in the Powell Group of the South Shetlands. Topographically, it is an ice-covered plateau rising to about 1200 or 1500 feet above sea-level. The rocks on the northern shore have been described by Mr. Wordie as contorted phyllites. The Lecturer's observations at Minstrel Bay on the western coast showed that the rocks there were similar. At Cape Lookout, on the south side of the island, a metamorphic series was encountered : this, according to Dr. C. E. Tilley, consists of amphibolite, garnet-albite-schist, quartz-hornblende-epidote-schist, and banded sandy limestone.

Observations from the ship were made of the volcanic island of Zavodovski, in the South Sandwich Group. The Tristan da Cunha Group in the Southern Atlantic, 1500 miles west of the Cape of Good Hope, was also visited. The islands are of volcanic origin. Particular attention was paid to the existence of Middle Island.

Gough Island lies more than 200 miles south of the Tristan da Cunha Group, and is 8 miles long by 3 miles in width. It is a monoclinal block, with dip-slopes to the west and escarpments to

the east. The lavas forming these features are basaltic, and intrusive into these lavas is a trachytic stock. Following this intrusion, the basalts were cut by a series of doleritic dykes. In general, it may be said that Gough Island presents many features similar to those that characterize the islands of Ascension and St. Helena.

February 16th, 1923.—Prof. A. C. Seward, Sc.D., F.R.S., President, in the Chair.

The President delivered his Anniversary Address on The Earlier Records of Plant-Life. Attention was drawn to the danger of excessive absorption in descriptive work, leading to insufficient consideration of such conclusions of general geological interest as can be drawn from the accumulated data. Reference was made to the views of Dr. Church on the origin of life in the waters of a primeval world-ocean, and on the origin of terrestrial vegetation from highly-organized Algæ transferred by emergence of portions of the Earth's crust above the surface of the water from an existence on the ocean-floor to life on land. It was suggested that the vegetation of the land may have received additions from upraised portions of the crust at more than one epoch in the history of the Earth. The course of evolution could probably be more correctly illustrated by the conception of separate lines of development, than by that of a branching tree implying the common origin of the main groups of plants. The unfolding of plant-life must be considered in relation to the changing geological background. The climatic and physical conditions of the Pre-Cambrian Era were briefly considered, and various kinds of indirect evidence of the existence of plant-life were critically examined: reference was made to graphite, supposed algal remains in association with oolitic structure, *Cryptozoon*, and the structures described by Dr. C. D. Walcott as Algæ or as the result of algal agency. Attention was called to the importance of carefully investigating diffusion-phenomena, as illustrated by the so-called Liesegang figures, as a possible explanation of the origin of some of the structures which are usually attributed to organic agency. We have no knowledge of any Pre-Cambrian land-flora. Palæobotanical records from Cambrian, Ordovician, and Silurian strata were briefly summarized, including some account of *Girvanella*, *Eophyton*, *Solenopora*, *Nematophycus*, *Pachytheca*, and *Parka*. Reasons were given for assigning some of the Cambrian Algæ described by Dr. Walcott to the Cyanophyceæ, especially *Marpolia spissa*.

In the second part of the Address, the older Devonian floras were reviewed, and some of the more characteristic genera described, special attention being directed to the petrified plants from the Rhynie chert-bed, discovered by Dr. W. Mackie, and described in detail by Dr. R. Kidston and Prof. W. H. Lang. Reference

was made to the differences between the older Devonian floras and those of Upper Devonian age. The question of the common origin of the phyla of Lycopods and Ferns was considered, and preference was expressed for the view which regards them as independently-evolved groups. In conclusion, the wide geographical range of *Archæopteris* was emphasized, and reference was made to the difficult problems raised by the occurrence of Upper Devonian floras well within the Arctic circle, at least equal (in the variety of the plants and in the vigorous development of the vegetation) to the more southern floras of Ireland, Belgium, and other regions.

February 28th, 1923.—Prof. A. C. Seward, Sc.D., F.R.S., President and, afterwards, Prof. W. W. Watts, Sc.D., F.R.S., Vice-President, in the Chair.

The following communications were read:—

1. 'The Late Glacial Stage of the Lea Valley (Third Report).' By Samuel Hazzledine Warren, F.G.S.

Since the publication of the previous papers on the subject, one new section of the same series of deposits has been found. This was in a different situation from the others, as it occurred at the level of, and in the area occupied by, the Middle or Taplow Terrace, whereas all the other sections were in the Low Terrace. It consisted of a bed of seed-bearing clay, in the middle of an old gravel-pit, partly built over, and consequently its precise stratigraphical relations to the Taplow gravel were not discoverable. The Taplow deposits yield a fairly temperate fauna and flora, and it is therefore concluded that the Arctic deposit cannot be of Taplow date. The site is close to the head of a small streamlet, and it is assumed, although it cannot be proved, that the Arctic plant-bed is of Low-Terrace or Ponders-End date, and that it represents the silting of a stream which flowed across the Taplow Terrace.

The paper is accompanied by a report on the Arctic flora by Mrs. E. M. Reid & Miss M. E. J. Chandler, in which some 48 species of flowering plants are recorded, and the conclusion is reached that there is nothing to distinguish the flora from that of the previously-described localities of the Lea Valley.

2. 'The *Elephas-antiquus* Bed of Clacton-on-Sea (Essex), and its Flora and Fauna.' By Samuel Hazzledine Warren, F.G.S.